

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2002, 01:33:20 ; Search time 1741.88 Seconds

(without alignments)  
17011.491 Million cell updates/sec

Title: US-09-357-675C-1

Perfect score: 1416

Sequence: 1 gccccactgcgtgcgcctnt.....aaaaaaaaaaaaaaaaaaaaa 1416

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sy:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: gb\_da:\*  
16: em\_fun:\*  
17: em\_fun:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_om:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_inv:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query	Score	Match length	ID	Description
-----					

1	1343	94.8	1385	9	AF069987	AF069987 Homo sapi
2	863.6	61.0	1365	10	BC021634	BC021634 Mus muscu
3	859.6	60.7	1338	10	AF069988	AF069988 Mus muscu
4	572.2	40.4	4079	9	AF069984	AF069984 Homo sapi
5	572.2	40.4	110000	2	AL554714_4	Continuation (5 of
6	572.2	40.4	110000	2	AL554714_5	Continuation (6 of
7	572.2	40.4	167863	2	AL590651	AL590651 Homo sapi
8	572.2	40.4	200882	9	AL591806	AL591806 Human DNA
9	356.2	25.2	1214	5	AF284575	AF284575 Xenopus l
10	326.8	23.1	4481	10	AF069985	AF069985 Mus muscu
11	326.8	23.1	211771	2	AC084821	AC084821 Mus muscu
12	326.8	23.1	213203	2	AC087229	AC087229 Mus muscu
13	308.4	21.8	178418	2	AC105589	AC105589 Rattus no
14	280	19.8	441	11	G13436	G13436 human SRS W
15	276.2	19.5	847	11	G72919	G72919 MARC 2849-2
16	182.8	12.9	1521	3	AF069989	AF069989 Drosophill
17	182.8	12.9	131853	2	AC017755	AC017755 Drosophill
18	182.8	12.9	168250	3	AC093121	AC093121 Drosophill
19	182.8	12.9	173494	3	AC093501	AC093501 Drosophill
20	182.8	12.9	174279	2	AC006169	AC006169 Drosophill
21	182.8	12.9	190351	3	AC105293	AC105293 Drosophill
22	182.8	12.9	298640	3	AE003467	AE003467 Drosophill
23	168.4	11.9	1385	3	AF069986	AF069986 Caenorhab
24	160.8	11.4	34013	2	AC010013	AC010013 Drosophill
25	144.2	10.2	1172	8	AF372904	AF372904 Arabidops
26	113.8	8.0	1063	10	BC020153	BC020153 Mus muscu
27	113.8	8.0	1292	10	AF284573	AF284573 Mus muscu
28	112.8	8.0	4762	1	AB017194	AB017194 Plectonem
29	112.6	8.0	25350	8	SPBC651	AL035570 S.pombe c
30	112.4	7.9	10466	1	AE004861	AE004861 Pseudomon
31	106.4	7.5	110908	1	SYCSLRD	D64002 Synchocyst
32	105.4	7.4	235050	1	AL414158	AL414158 Yersinia
33	105.2	7.4	203050	1	AL646071	AL646071 Ralstonia
34	104.4	7.4	169364	3	CEY56A3A	AL132860 Caenorhab
35	102.6	7.2	343550	1	AP003587	AP003587 Nostoc sp
36	102	7.2	962	9	AF284574	AF284574 Homo sapi
37	102	7.2	1019	9	AF260334	AF260334 Homo sapi
38	101.4	7.2	10765	1	AE008331	AE008331 Agrobacte
39	101.4	7.2	11689	1	AE009281	AE009281 Agrobacte
40	101.2	7.1	57093	2	AC012711	AC012711 Drosophill
41	101.2	7.1	148847	3	AC008234	AC008234 Drosophill
42	101.2	7.1	175053	3	AC010581	AC010581 Drosophill
43	101.2	7.1	224400	3	AE003682	AE003682 Drosophill
44	100.4	7.1	965	9	BC020620	BC020620 Homo sapi
45	99.8	7.0	12748	1	AE005759	AE005759 Caulobact

## ALIGNMENTS

RESULT	1	AF069987	1385 bp	mrna	linear	PRI 23-JUL-1998
LOCUS	AF069987	Homo sapiens	nitrilase 1 (NIT1)	mrna	complete cds.	
DEFINITION	AF069987	AF069987				
ACCESSION	AF069987	GI:3228665				
VERSION	AF069987.1	GI:3228665				
KEYWORDS						
SOURCE						
ORGANISM						
		human.				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE		1 (bases 1 to 1385)				
AUTHORS		Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,				
		Tilib,S., Draganesu,A., Wermuth,P., Rothman,D.H., Huebner,K.,				
		Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.				
		Nitrilase and phl homologs are encoded as fusion proteins in				
		Drosophila melanogaster and Caenorhabditis elegans				
		Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)				
TITLE		2 (bases 1 to 1385)				
JOURNAL		Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,				
MEDLINE		Tilib,S., Draganesu,A., Wermuth,P., Rothman,D., Huebner,K.,				
REFERENCE		Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.				
AUTHORS		98337986				

TITLE Direct Submission  
JOURNAL Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Ninth St., Philadelphia, PA 19107, USA  
FEATURES  
source  
1.1385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q21-q22"  
1.1385  
/gene="NIT1"  
77.1060  
/gene="NIT1"  
/note="NIT1"  
/codon\_start=1  
/product="nitric oxide synthase 1"  
/protein\_id="AAC39907.1"  
/db\_xref="GI:3228666"  
/translation="MUGFTTRPPHRTSLICPGRLIPQLSVLCAQPPRAMAIISSSC  
ELPLAVCOVYSTPDKQNFRTCAELVREARLQACLAFLPEADFLARDPAETLHS  
EPLGKLEBETQLAREGLMLSLGDFHERGQDWEQKLYNCHVFLNSKGAAYATR  
KTHLCDEIVEPGQPMCESTMPGSLSESPVPAKIGLAVCYDMPEPELSIALAQA  
GAEIIVPSAFSGSITGPAMHVEVLRLARAITOCYVVAACGRRHERRASGHSMMVD  
PMGVNVARCSBEGELCLARDLNTLRRLRHLRPFQHRREDLIGNLHPLS"  
BASE COUNT 322 a 382 c 357 g 324 t  
ORIGIN  
Query Match 94.8% Score 1343; DB 9; Length 1385;  
Best Local Similarity 97.7%; Pred. No. 0;  
Matches 1384; Conservative 0; Mismatches 1; Indels 31; Gaps 1;  
QY 1 gccacatcgctgctgctccttgcctcagacacgcccctcgagatcgacccctgcaatg 60  
Db 1 GCCCACTCGCTGCGGCTATCTGGCTCCAGACCGCCCTCGGATCGGACCCGGGAATGG 60  
QY 61 tttagctatctctcatgtagaactactccatcccgctcgccgctgagctgctca 120  
Db 61 TTTTGGCTATATCTTTCAT-----GCTGGGCTTCA 89  
QY 121 tcaccagagccctcagacgattcctgtccctctgtctcctgagctcggaactcaac 180  
Db 90 TCACCAAGGCTCTCCACAGATTCTGTCTCTGTCTGTGACTCCGATACCTCCAC 149  
QY 181 tctcagtaactgtgtcctcagccagccagacagctatctcctcctcctcctg 240  
Db 150 TCTCAGTACTTGTGTCTGACGCCAGGCCAGACGATGCTCTCTCTCTCTCTG 209  
QY 241 aactgcccctgctgctgtgtcagagtaacatcgacgacgacgaagaacaaactta 300  
Db 210 AACTGCCCTGT 269  
QY 301 aaacatgtcctgagctgtgtcagagctgtcagacgctgagctgtcctgcttctgc 360  
Db 270 AAACATGTGCTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 329  
QY 361 ctgagagcatttgactatcgtcagcaggaacccctgagagcactaaccctgctgaacc 420  
Db 330 CTGAGGCAATTGACTTATGTGACAGGACCCCTGCAGAGACCTTACCTGTCTAACCAC 389  
QY 421 tgggtgggaaactcttggaagaatacaccagctgtccagggaaatgtggaactgtgct 480  
Db 390 TGGGTGGGAACCTTTTGAAGAATACACCCAGCTTGCAGAGGAATGTGAGACTGTG 449  
QY 481 cctgggtgtgttcacatgagcgtgtgccaagactgtggagcagacccagaatactacaatt 540  
Db 450 CCTTGGGTGTGTTCATGAGAGGTGTGCCAAGACTGTGGAGCACTCAAGAAATCTAACAAT 509  
QY 541 gtcacgtgctgtcgaagaagaagaagggcagtagtgcgaactacaggaagaacatctgt 600  
Db 510 GTACACGTCGTGTGACAGCAAGGAGGAGGAGTGTGAGGACCTTACAGGAGACATCTGT 569  
QY 601 gtgacgtagagatccagggcagggcctatgtgtgaaagcaactctacatgcttggc 660

Db 570 GTGACGTAGATTCACAGGAGGAGGCTATGTGAAAGCACTACATGCTGGGC 629  
QY 661 ccaagctgagatccactctcagacacacagaggaagaataatgtcctcctcctg 720  
Db 630 CCAGCTTGAGTACCTCTCAGCACACAGGAGGAAATGTTGTACTGTCTGTG 689  
QY 721 acatgagttccctgaactctcctcctgcatgtgtcctcagctgaggaagaatactact 780  
Db 690 ACATGCGGTTCCCTGAATCTCTGTGCGATTGGCTCAAGCTGAGACAGATCTACT 749  
QY 781 atcctcagcttcttgatcctatcagagccagccactgagggagtgctgctgggccc 840  
Db 750 ATCTTCAGCTTTTGGATCCATTACAGGCCACGACCTGAGAGTGTGTGCGGGGCC 809  
QY 841 gtgctatgaaacccagctgctatgtgtgagagagagcagtgatggaagcaccatgaga 900  
Db 810 GTGCTATGAAACCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 869  
QY 901 agagagcaagtatggtccacagacatgtgtgtgagacccctgagggaacagtggtggcc 960  
Db 870 AGAGAGCAAGTATGAGCCACAGACATGTTGTGTGTGTGTGTGTGTGTGTGTGTGT 929  
QY 961 gctctgagggcagggcctctgcttgccttgcggaatagacctaactatctgcagtgct 1020  
Db 930 GCTCTGAGGGGCCAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 989  
QY 1021 gccgaacactgctgtgtgttcagacacgcaagcctgacactatggaatctggtgacc 1080  
Db 990 GCCGACACCTGCTGT 1049  
QY 1081 cactgtcttaagactgactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140  
Db 1050 CACTGTCTTAAGACTGTGACTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1109  
QY 1141 actatgagtagtgctcatgtgactgtgaggtgaggtgaggtgaggtgaggtgaggtg 1200  
Db 1110 ACTATGAGCTGT 1169  
QY 1201 gagaacactgactctctgt 1260  
Db 1170 GAGAACCTTGTGCTCTGT 1229  
QY 1261 gagctcaccctgaggtcagactgtcaggtcaggtcaggtcaggtcaggtcaggtcag 1320  
Db 1230 GAGCTTACCTGAGGTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGAC 1289  
QY 1321 tatttcataagactgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380  
Db 1290 TATTTCATGGAACCTGAAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1349  
QY 1381 taatcataaagtcataaagtcataaagtcataaagtcataaagtcataaagtcataa 1446  
Db 1350 TAAATCATTAAGTCAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1385  
RESULT 2  
BC021634 1365 bp mRNA linear ROD 22-JAN-2002  
LOCUS BC021634  
DEFINITION Mus musculus, nitric oxide synthase 1, clone MGC:13825 IMAGE:4008543, mRNA,  
complete cds.  
ACCESSION BC021634  
VERSION BC021634.1 GI:18204912  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1365)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer



SOURCE	Mus musculus.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 1338)
AUTHORS	Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tillib, S., Draganescu, A., Wernuth, P., Rothman, J.H., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.
TITLE	Nitrlase and flhl homologs are encoded as fusion proteins in drosophila melanogaster and caenorhabditis elegans
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
MEDLINE	98337986
REFERENCE	2 (bases 1 to 1338)
AUTHORS	Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tillib, S., Draganescu, A., Wernuth, P., Rothman, J., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.
TITLE	Direct Submission
JOURNAL	Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES	Location/Qualifiers
SOURCE	1..1338
gene	/organism="Mus musculus"
CDS	/db_xref="taxon:10090"
	/chromosome="1"
	/map="1q21-q23"
	1..1338
	/gene="Ntll"
	58..1029
	/codon_start=1
	/product="nitrlase 1"
	/protein_id="AAC0185.1"
	/db_xref="GI:3228668"
	/translation="MLGIFTRPPHQLCTGYRLRTPVLICTOPPRMSSSTWELPPIGVAVOCVSPMKOEKFEIETLVEARILAGLCALPEAPFIRNAEITLLSEPLINDLGIOVSOLAREGGIWLSIGSFHEKODMNQKITYNCHILSKSSVANYSYKTHCDVIPROGPMBRESNTYKPGTLEPPVKTRPGAKGLAICIDMRPELSLKQAIAELTYTSAFGSVTGPAHWELTLVARIESQCXYIAAOGRHHTRAFSGHSMVPDPMGTWVASCSGPGICLARIDLHFHPQQRORHLPVQHRRPDLXSLGHPLS"
BASE COUNT	347 a 335 c 330 g 326 t
ORIGIN	
Query Match	60.7%; Score 859.6; DB 10; Length 1338;
Best Local Similarity	82.1%; Pred. No. 1.4e-242;
Matches 1072; Conservative	0; Mismatches 204; Indels 30; Gaps 6;
OY	110 gctgggcttcacccagcctccccaagaatctctgtccctctgtgtcctgaactcg 169 
Db	60 GCTGGGCTTCACACCACCAAGCCCTCCACCAA-----CTCCTGTATACCGGATACCG 110 
OY	170 gatacctaacctcagtaactttgtctgcagaccagccagccagagcatctctc 229 
Db	111 ATTACTTGGAACCCAGACTTTGTACTACGAGCCAGGCCAGAACCATGTC---CTCATC 167 
OY	230 ttccctctcgaaactgcgcccttggtgctgtgtctgcagglaacaatcgagccagaacaqa 289 
Db	168 AACCTCCCTGGGAGCGCCCTCTGTGGCTGTGNGCCAGTAGCATCAACACCAACAACAGA 227 
OY	290 acagaacttaaaaatatgtctgtagctgtgttcgagaagctgccagaactggagtctgct 349 
Db	228 AGAGAACTTTAAAACATGTGCTGACTTTGGTTCCAAGAGCTGCACATCGGGGTCTTGCT 287 
OY	350 gagcttctgctggagagatttgaacttcatacgcaaggagaccttgagagaagctcacact 409 
Db	288 GGCTTTTGCTGCTGAGGCAATTGGACTTTATTTGACAGAAACCCCTGCCGAGACATTTCCT 347 
OY	410 gtctgaaccactgggttggaactctlttgagaagaatataccaccagcttgcagaggaatftg 469 
Db	348 GTCCGAACCACTGATGGAGATCTTTTGGGCAATAATATAGCCAGCTTGGCAGGAGATGTCG 407 
OY	470 acctgtgcttctcttgtgtgttccatgatgacgtgtggccaagacttggagaggaactcagaa 529 

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE
AF069984	4079 bp DNA	linear	PRI 23-JUL-1998		human.	Homo sapiens
LOCUS	Homo sapiens ntr1riase homolog 1 (NIT1) gene, alternatively spliced	AF069984				
DEFINITION	product, complete cds.	AF069984				
ACCESSION		AF069984.1	GI:3242977			
VERSION						
KEYWORDS						
ORGANISM	human.					
SOURCE	Homo sapiens					





QY 869 ggcacacacagctgtgagccaccacatgagagagcaagttatggccacagcatgt 928  
 |||||  
 Db 105614 GGCACACACAGCTGTGGAGCGCCACCATGAGAGAGCAATTTATGGCCACGATGGT 105673  
 QY 929 gtagaacccctggggaacagtggtggccgcgtctgtctgaggggccaagcctctgtctgc 988  
 |||||  
 Db 105674 GGTAGACCCCTGGGGAACAGTGTGGCCGCGTCTGTGAGGGGCGAGGCGCTGTGCCTTGC 105733  
 QY 989 ccgaatagacctcaactatctgagcagttgtggccgaacacgcctgtgttccagacag 1048  
 |||||  
 Db 105734 CCGAATTAACCTCAACTATCTGCGAGTGTGGCGGACACCTGCTGTGTTCCACACCG 105793  
 QY 1049 caggcctacactctatgtgcaatctgtgttaaccactgtctttaaagcttgaactgtgag 1108  
 |||||  
 Db 105794 CAGGCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGCTTGACTTCTGTGAG 105853  
 QY 1109 tttagacctgcccccccaaccccccaacctgccaactatgagctagtctcatgttgaactgg 1168  
 |||||  
 Db 105854 TTTAGACCTGGCCCTCCACCCCGCCACCTGCACATGAGTATGCTCATGTGACTTGG 105913  
 QY 1169 aggcagatccagagacagctctccctcaactgtgagaaaccttgacctctctgtatgaaacac 1228  
 |||||  
 Db 105914 AGGCAGATCCAGGACACAGCTCCCTCACTTGGAGAACCTTGACTCTTGTATGGAACAC 105973  
 QY 1229 agatggctgtctgggaaagaaccttcaacctgagcttcaacctgaggtcaagctgaagt 1288  
 |||||  
 Db 105974 AGATGGCTGCTGTGGGAAGAAACTTTCACCTGACCTTACCTGAGGTGACAGTGCAGTT 106033  
 QY 1289 tcagaagaagtggaatttatatagtcattgttatttcaatggaacctgaagttctgtcgtga 1348  
 |||||  
 Db 106034 TCAGAAGAAGTGAATTTATATAGTCATGTTATTTTCATGGAACCTGAAGTTCTGCTGA 106093  
 QY 1349 gggctgagcagcagctggtcattgaaaaataataataatcaataagtc 1393  
 |||||  
 Db 106094 GGGCTGAGCAGCAGTGGCAATGAAAAATATATATATCAATAAAGTC 106138

RESULT 6  
 AL354714.5  
 WPCOMMENT

Sequence split into 7 fragments LOCUS AL354714 Accession AL354714  
 Fragment Name Begin End  
 AL354714\_0 1 110000  
 AL354714\_1 100001 210000  
 AL354714\_2 200001 310000  
 AL354714\_3 300001 410000  
 AL354714\_4 400001 510000  
 AL354714\_5 500001 610000  
 AL354714\_6 600001 699820  
 Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome 1 c

Query Match 40.4%; Score 572.2; DB 2; Length 110000;  
 Best Local Similarity 98.6%; Pred. No. 2.9e-157;  
 Matches 577; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 809 cccaaccacactggaggggtgtgtctgtggccggtgtctatcgaaacccagtgctatgtagt 868  
 |||||  
 Db 5554 CACATCTCATGCGCCAGGTTGCTGTCGGGCCGCTCTATCGAACCACACTGATGTAGT 5613  
 QY 869 ggcacacacagctgtgagccaccacatgagagagcaagttatggccacagcatgt 928  
 |||||  
 Db 5614 GGCACACACAGTGTGGAGCGCCACCATGAGAGAGCAATTTATGGCCACGATGGT 5673  
 QY 929 gtagaacccctggggaacagtggtggccgcgtctgtctgaggggccaagcctctgtctgc 988  
 |||||  
 Db 5674 GGTAGACCCCTGGGGAACAGTGTGGCCGCGTCTGTGAGGGGCGAGGCGCTGTGCCTTGC 5733  
 QY 989 ccgaatagacctcaactatctgagcagttgtggccgaacacgcctgtgttccagacag 1048  
 |||||  
 Db 5734 CCGAATTAACCTCAACTATCTGCGAGTGTGGCGGACACCTGCTGTGTTCCACACCG 5793

QY 1049 caggcctacactctatgtgcaatctgtgttcaaccacactgtctttaaagcttgaactgtgag 1108  
 |||||  
 Db 5794 CAGGCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAGCTTGACTTCTGTGAG 5853  
 QY 1109 tttagacctgcccccccaaccccccaacctgccaactatgagctagtgtctcatgttgaactgg 1168  
 |||||  
 Db 5854 TTTAGACCTGGCCCTCCACCCCGCCACCTGCACATGAGTATGCTCATGTGACTTGG 5913  
 QY 1169 aggcagatccagagacagctctccctcaactgtgagaaaccttgacctctctgtatgaaacac 1228  
 |||||  
 Db 5914 AGGCAGATCCAGGACACAGCTCCCTCACTTGGAGAACCTTGACTCTTGTATGGAACAC 5973  
 QY 1229 agatggctgtctgggaaagaaccttcaacctgagcttcaacctgaggtcaagctgaagt 1288  
 |||||  
 Db 5974 AGATGGCTGCTGTGGGAAGAAACTTTCACCTGACCTTACCTGAGGTGACAGTGCAGTT 6033  
 QY 1289 tcagaagaagtggaatttatatagtcattgttatttcaatggaacctgaagttctgtcgtga 1348  
 |||||  
 Db 6034 TCAGAAGAAGTGAATTTATATAGTCATGTTATTTTCATGGAACCTGAAGTTCTGCTGA 6093  
 QY 1349 gggctgagcagcagctggtcattgaaaaataataataatcaataagtc 1393  
 |||||  
 Db 6094 GGGCTGAGCAGCAGTGGCAATGAAAAATATATATCAATAAAGTC 6138

RESULT 7  
 AL590651/c 167863 bp DNA linear HTG 18-AUG-2001  
 LOCUS Homo sapiens chromosome 1 clone RP11-137A12, \*\*\* SEQUENCING IN  
 DEFINITION  
 PROGRESS \*\*\* 9 unordered pieces.  
 ACCESSION  
 AL590651  
 AL590651.4 GI:13990236  
 VERSION  
 HTG: HTGS\_PHASE1; HTGS\_CANCELLED.  
 KEYWORDS  
 human.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 Requests: clonerequest@sanger.ac.uk  
 On May 7, 2001 this sequence version replaced gi:13751001.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 Project Information  
 Center project name: ba137A12  
 ----- Summary Statistics  
 Assembly program: XGAP; version 4.5  
 Sequencing vector: Plasmid; 108752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 164729 bases at least Q40  
 Consensus quality: 165486 bases at least Q30  
 Consensus quality: 166169 bases at least Q20  
 Insert size: 167063; sum-of-contigs  
 Insert size: 171878; 3.4% error; agrose-fp  
 Quality coverage: 8.28x in Q20 bases; sum-of-contigs quality  
 coverage: 8.11x in Q20 bases; agrose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 9373: contig of 9373 bp in length  
 \* 9374 9473: gap of 100 bp

QY	989	ccgataagactcaactctcgcgaacgttcgccgaacaacctgcttgcttccaacaccy	1048
Db	149516	CGGATTAACCTCAACTAATTGCAGCAACTTGGCGCACAACCTGCTGTGTTCCACACC	149457
QY	1049	caggcctcaactctataggcaatctcggfscaccocactgtctctaagaacttgactctgttag	1108
Db	149456	CAGGCTTAACTCTATGGCANCTTCGGTCAACCACTGTCTTAAGACTTGACTTCGTAG	149397
QY	1109	ttagaacctgccectcccacccccctcgcoactatagagtatgtctcatgttaactggy	1168
Db	149396	TTTTAGACCTGCCCTCCACCCCCCACCTGCCACTATAGCTAGTCAATGCTCATGACTTGG	149337
QY	1169	aggcagaagttccaagcacagcgtcccccactctggggaaccttgactctttgatatgsaaac	1228
Db	149336	AGGCAGAGTTCACAGGCACGCTCCCTCCATCTGGAGAACCCTTCACCTCTGTGAAGAACAC	149277
QY	1229	agatggactgtcttggaaagaaccttcaacctgaagcttcaacctgagtgfccaagtcaatt	1288
Db	149276	AGATGGCTGCTTGGGAAGAACCTTCACTGAGCTTCACCTGAGGTGCAGACTCACTT	149217
QY	1289	tcgaagaagtgtaaatltatatagtcacatgttatlcatlgaagaaactgaagtlctgtcga	1348
Db	149216	TCAGAAGCTGCAATTTATATAGCATTTGTTATTTTCATGGAACCTGAAGTTCGTGCA	149157
QY	1349	gggctgagcagcactgycatlgaaaataataataatcataaagtc	1393
Db	149156	GGGCTGACACACACTGCGCATTGAAAAATAATATATCAATAAAGTC	149112
RESULT	8		
ALYS1806		20822 bp	DNA linear PRI 30-JAN-2002
LOCUS	AL591806		
DEFINITION	Human DNA sequence from clone Rpl1-544M22 on chromosome 1, complete		
ACCESSION	AL591806		
VERSION	AL591806.16	GI:18476709	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 20822)		
AUTHORS	Harrison,E.		
JOURNAL	Direct Submission Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 1, 2002 this sequence version replaced gi:17902927. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.		
COMMENT	This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WormPEP; Information on the WormPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence was generated from part of bacterial clone conligs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <a href="http://www.sanger.ac.uk/HGP/ChrlRpl1-544M22">http://www.sanger.ac.uk/HGP/ChrlRpl1-544M22</a> is from the library RPCL-11.2 constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a>		



[illegible]

Query Match	Best Local Similarity	23.1%	Score 326.8	DB 10	Length 4481
Matches 466; Conservative	77.7%	Pred. No. 5.5e-85	Mismatches 112	Indels 22	Gaps 5
Db 801	attacagagccagcccaactgagagtgctgctgagcccgctgctatcgaaaccagctgc	860			
Db 3368	AGTCACGACGCGTCATCTCAGAGTGCTGCGGGCCGCCCATTTGATCTCAGATGC	3427			
QY 861	tatgtatgtagtcagcagcagcagctgtgagcgcacacatgagaagagcaagtattgcccac	920			
Db 3428	TATGTATATAGCACACGCGCGTGTGACGCCACCATTAACAAACAGCAAGTTATGGCCAT	3487			
QY 921	agcattgtgttagaaccccttgagaaacagtggtgagcccgctgctctgaggggcagagctc	980			
Db 3488	AGCATGTGTGTGACCCGTGTGGGCACACGTGTGGCCCGCTGCTCCGAGGGACCGGCTC	3547			
QY 981	tgccttgcacgaatagacctcaactatctcgacagctgtgcgcgacacactgtcctgtgtc	1040			
Db 3548	TGCGTTGCTCGAATTGATCTCCACTTTCTACAAACAGATGGCGCAACGCTGCTGTTC	3607			
QY 1041	cagaccgcagagccttgcactctatggcaatctgggtcaccacactgtcttaaac--ttga	1098			









RESULT	14
G13436/c	
LOCUS	G13436
DEFINITION	human STS WI-12338, sequence tagged site.
ACCESSION	G13436
VERSION	G13436.1 GI:1127545
KEYWORDS	STS; STS sequence; primer; sequence tagged site.
SOURCE	human STS derived from sequences in dbEST and the Unigene collection.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 441)
JOURNAL	Hudson, T.
COMMENT	Whitehead Institute/MIT Center for Genome Research; Physically mapped STS Unpublished

Buffer:	MgCl2: 1.5 mM
	KCl: 50 mM
	Tris-HCl: 10 mM
	pH: 9.3
FEATURES	Derived from dbEST (genbank accession T78621).
source	Location/Qualifiers
	1..441
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/map="750.8 CR from top of Chr1 linkage group"
STS	16..145
primer_bind	16..38
primer_bind	complement(124..145)
BASE COUNT	107 a 95 c 130 g 103 t 6 others
ORIGIN	

Query Match	19.8%	Score 280	DB 11	Length 441
Best Local Similarity	98.4%	Pred. No. 2,4e-71		
Matches 313	Conservative	0	Mismatches 2	Indels 3
QY 1079	cccaactgcttaagaacttga-cttcgtgtgagttagaacctgcccctccaccaccaccct	1137		
Db 317	cccaactgcttaaanacttgacctcttgagagtttagacctggccctccaccaccaccct	258		
QY 1138	gccaa-ctattgagctagtgtcctaagtgtgaacttggagagcaggaaccagggcacagcttcccctta	1196		
Db 257	gccaaactgtgagcctgagctctcatgtgacctggagagcagagancagagcagcagctcccctca	198		
QY 1197	cttggagaacctgtgctctctcttggatggagaaacaagaattggctctgtcttgggaaaaaactttc	1256		
Db 197	cttgagagaaacctgtgactctcttggatggagaaacaagatgggctcttgggaaaaaacttttc	138		
QY 1257	acctgagcttaaccctgagagctcagactgcagtttcagaaagagtggaattatataagcat	1316		
Db 137	acctgagcttcaaccctgagagctcagactgcagtttcagaaagagtggaattatataagcat	78		
QY 1317	tgattatttcaltgysaaactgaaagctctgctgagggctgagcagcaacttgcattgaaaat	1376		
Db 77	tgtttattttcatggaacatgaaagttctgtctga-ggctgagcagcactgcattgaaaat	19		
QY 1377	ataataatcataaagtcata	1394		
Db 18	ataataatcataaagtcata	1		
RESULT 15				
LOCUS G72919				
DEFINITION	G72919	847 bp	DNA	linear
ACCESSION	MA6C 2849-2850:99193517:1	SCF	porcine spleen	Sus scrofa STS
VERSION	G72919			
KEYWORDS	G72919.1	GI:15146949		
SOURCE	STS			
ORGANISM	pig			
REFERENCE	Sus scrofa			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.			
	1 (bases 1 to 847)			
	Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and			
	Keele,J.W.			
	Single nucleotide polymorphism (SNP) discovery in expressed porcine			
	genes			
	Unpublished (2001)			
JOURNAL COMMENT				
	Contact: Freking BA			
	USDA, ARS, US Meat Animal Research Center			
	PO Box 166, Clay Center, NE 68933-0166, USA			
	Tel: 402 762 4278			
	Fax: 402 762 4173			
	Email: freking@mail.marc.usda.gov			



---

THIS PAGE BLANK (USPTO)



XX 20-JUL-1998; 9805-0093350.  
PR XX  
XX  
PA (UYJE-) UNIV JEFFERSON THOMAS.  
XX  
XX  
PI Croce CM;  
XX  
XX  
DR WPI: 2000-171195/15.  
DR P-PSDB; MAY68739.  
XX  
XX Novel nitrilase homologs used as diagnostic and therapeutic reagents  
PT for the detection and treatment of cancer -  
XX  
XX Claim 6; Fig 6; 25pp; English.  
XX

CC The present sequence represents the coding region of human, murine,  
CC *Drosophila melanogaster* and *Caenorhabditis elegans* NIT1 gene. The  
CC human and mouse NIT1 genes are members of an uncharacterised  
CC mammalian gene family with homology to bacterial and plant nitrilases.  
CC The tumour suppressor gene FHIT in *D. melanogaster* and *C. elegans* codes  
CC for fusion proteins in which the Phit domain is fused with a Nit domain.  
CC In mouse and humans, FHIT and NIT are encoded by two different genes,  
CC localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The  
CC human FHIT gene at chromosome 3p14.2, spanning the constitutive  
CC chromosomal fragile site FRAB, is often altered in most common forms  
CC of human cancer. The Nit1 protein overcomes the mutated inactivation  
CC of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives  
CC and analogues of them, and antibodies are used as diagnostic and  
CC therapeutic reagents for the detection and treatment of cancers.

Query Match	99.7%	Score 1411.8;	DB 21;	Length 1416;
Best Local Similarity	99.98%;	Pred. No. 0;		
Matches 1414; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

OY	1	gccacatcgctgcggcgcctnctctgtgcttcacagacgcgcctccggaatcgagaccctgcgaatg	60
Db	1	gccacatcgctgcggcgcctnctctgtgcttcacagacgcgcctccggaatcgagaccctgcgaatg	60
OY	61	ttttgacatactctcatgtctgagaccctaccccatccgctcgagccggctggtgattca	120
Db	61	ttttgacatactctcatgtctgagaccctaccccatccgctcgagccggctggtgattca	120
OY	121	tcaccagagcctctctcaagaattcctgtccctctctgtctgtgacttcggatacctcaac	180
Db	121	tcaccagagcctctctcaagaattcctgtccctctctgtctgtgacttcggatacctcaac	180
OY	181	tctcagatccttctgtctcgaagccgagccgagagcagcatctatcctctcctctctgg	240
Db	181	tctcagatccttctgtctcgaagccgagccgagagcagcatctatcctctcctctctgg	240
OY	241	aactgcgccctctgtgctgtctgtgcagagtaacatctgacgcgcagacagcaagcaagaactta	300
Db	241	aactgcgccctctgtgctgtctgtgcagagtaacatctgacgcgcagacagcaagaactta	300
OY	301	aaacatgtctgcagacctgtcttcagagagctgcacagactctgagctgtgctctctctgc	360
Db	301	aaacatgtctgcagacctgtcttcagagagctgcacagactctgagctgtgctctctctgc	360
OY	361	ctgagagcatttgaacttcattctgcacgaggaccctgcagagacgctacacactctgaaaccac	420
Db	361	ctgagagcatttgaacttcattctgcacgaggaccctgcagagacgctacacactctgaaaccac	420
OY	421	ctgggtgaggaacacttttggaagaatacagccacgcctctgcagaggaattgtgacctctgctt	480
Db	421	ctgggtgaggaacacttttggaagaatacagccacgcctctgcagaggaattgtgacctctgctt	480
OY	481	cctctgtgtgttcttcacatgagctgtgcacagactctggagcagactcagaaatctacaatt	540
Db	481	cctctgtgtgttcttcacatgagctgtgcacagactctggagcagactcagaaatctacaatt	540

QY	54.1	gtctcgtgcgtctgtgaacacgaacgaagggccaaatgaatggccacttaccgaagaagacacatcgtt	600
Db	54.1	gtcaacgtgcgtctgtgaacacgaacgaagggccaaatgaatggccacttaccgaagaagacacatcgtt	600
QY	60.1	gtgaaagctgaagatctccaaaggccaaagggtgcctatgtgtgtaaagaacactctaccacgtgcctgggc	660
Db	60.1	gtgaaagctgaagatctccaaaggccaaagggtgcctatgtgtgtaaagaacactctaccacgtgcctgggc	660
QY	66.1	ccaagctctgaagtcacccgtctcaagcaacacaaagcagaagatgtgtctcaagctgtctgtcatag	720
Db	66.1	ccaagctctgaagtcacccgtctcaagcaacacaaagcagaagatgtgtctcaagctgtctgtcatag	720
QY	72.1	acaatgcggtcttcccgaaactcctctctgtgcattctgaagctcgaagcttgaaagagaataactaacct	780
Db	72.1	acaatgcggtcttcccgaaactcctctctgtgcattctgaagctcgaagcttgaaagagaataactaacct	780
QY	78.1	atcccttcagccttttltgatalccatlttaacaaagcccaagcccaactctgtagagatgtgtctgcgtgcgcc	840
Db	78.1	atcccttcagccttttltgatalccatlttaacaaagcccaagcccaactctgtagagatgtgtctgcgtgcgcc	840
QY	84.1	gtgcctatctgaaaccccaagctgtcatgtatgtatgtgcagaagacaagatgtgtgaaagcccaacatagaa	900
Db	84.1	gtgcctatctgaaaccccaagctgtcatgtatgtatgtgcagaagacaagatgtgtgaaagcccaacatagaa	900
QY	90.1	aagaaagcaagatctatgtgcacacagacatgtgtgtgtaaagccctctgggtgaaagaagatgtgtcccggt	960
Db	90.1	aagaaagcaagatctatgtgcacacagacatgtgtgtgtaaagccctctgggtgaaagaagatgtgtcccggt	960
QY	96.1	gctctcgaaagggccaaagcgcctctgcctctgtcccggaatagaactccaactatctctgcgaacagtgtgc	1020
Db	96.1	gctctcgaaagggccaaagcgcctctgcctctgtcccggaatagaactccaactatctctgcgaacagtgtgc	1020
QY	102.1	gccgaacacctcgcgtgtgtcttcacagaacccgaagcctctgaacctctatgtgcaactctgtgtcacac	1080
Db	102.1	gccgaacacctcgcgtgtgtcttcacagaacccgaagcctctgaacctctatgtgcaactctgtgtcacac	1080
QY	108.1	caatgtctcttaagacttgaactctgtgtgaagtlttaagactcgtccctccaccacccccaacctgtgc	1144
Db	108.1	caatgtctcttaagacttgaactctgtgtgaagtlttaagactcgtccctccaccacccccaacctgtgc	1144
QY	114.1	actatgaagctgtgtcctatataatgtatgtatgtgaagacagagatcccaagacagagcttcccctcaactgt	1200
Db	114.1	actatgaagctgtgtcctatataatgtatgtatgtgaagacagagatcccaagacagagcttcccctcaactgt	1200
QY	120.1	gagaacacctctgaactcctctgtatgtgaaacacaaatagtggtcgtgtctgtggaagaagaacttttaacct	1266
Db	120.1	gagaacacctctgaactcctctgtatgtgaaacacaaatagtggtcgtgtctgtggaagaagaacttttaacct	1266
QY	126.1	gaagcttcaacctgtgaaggttcaagactgtgaaggttcaagaagaaggtgtgaatttataatagaatcattgtt	1322
Db	126.1	gaagcttcaacctgtgaaggttcaagactgtgaaggttcaagaagaaggtgtgaatttataatagaatcattgtt	1322
QY	132.1	taattctacgtgaagaactcgtgaatctgtctgtgaaggtcgtgaagcaagctgcgtcatgtgaaataatata	1381
Db	132.1	taattctacgtgaagaactcgtgaatctgtctgtgaaggtcgtgaagcaagctgcgtcatgtgaaataatata	1381
QY	138.1	taatcataaagtccaataaataaataaataaataaataa 1416	
Db	138.1	taatcataaagtccaataaataaataaataaataaataa 1416	

RESULT	2
AAFI6257	
ID AAFI6257	standard; cDNA; 1382 BP.

AA AF16257

DT 13-MAR-2001 (first entry)

AA	DE	Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
AA	DE	Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.

AA Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;

KW vulnery; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease; ss.  
XX Homo sapiens.  
XX MO200055174-A1.  
XX PD 21-SEP-2000.  
XX PD 08-MAR-2000; 2000WO-US05988.  
XX PE 12-MAR-1999; 99US-0124270.  
XX PR (HUMA-) HUMAN GENOME SCI INC.  
XX PA (ROSE/) ROSEN C A.  
XX PI Rosen CA, Ruben SM;  
XX DR WPI: 2000-587513/55.  
XX DR P-PSDB: AAB57054.  
XX PT Prostate cancer associated gene sequences, referred to as prostate  
XX PT cancer antigens, useful for treatment, prevention, and diagnosis of  
XX PT disorders such as prostate cancer -  
XX PS Claim 1; Page 1124; 2338pp; English.  
XX CC AAF15566 to AAF16505 encode the human prostate cancer associated  
XX CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
XX CC The prostate cancer antigens can have neuroprotective, cytosstatic,  
XX CC cardiactive, immunomodulatory, muscular, vulnery, gastrointestinal,  
XX CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
XX CC and can be used in gene therapy. The prostate cancer antigen  
XX CC polynucleotides may be used for detection of prostate cancer, chromosome  
XX CC identification, as chromosome markers, and for numerous other diagnostic  
XX CC or research purposes. The prostate cancer antigens may be used to treat  
XX CC disorders such as neural, immune, muscular, reproductive,  
XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
XX CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
XX CC AAB57303 represent sequences used in the exemplification of the present  
XX CC invention.  
XX CC  
XX CC Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other:  
SQ

Query Match 93.3%; Score 1321; DB 21; Length 1382;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 1376; Conservative 0; Mismatches 6; Indels 32; Gaps 2;

QY 1 gccacatgctgctgctcttctgctcagacgcctcctcgatcgagcctcgaaatg 60  
DB 1 gccacatgctgctgctcttctgctcagacgcctcctcgatcgagcctcgaaatg 60  
QY 61 ttgttgcatatcttcagatgtagactactccctaccgtgcgcgcgtcggttca 120  
DB 61 ttgttgcatatcttcagatgtagactactccctaccgtgcgcgcgtcggttca 120  
QY 61 ttgttgcatatcttcagatgtagactactccctaccgtgcgcgcgtcggttca 120  
DB 61 ttgttgcatatcttcagatgtagactactccctaccgtgcgcgcgtcggttca 120  
QY 121 tcacacagcctcctcacagatctcctcctctgtctcgtacgtcgatccgaac 180  
DB 90 tcacacagcctcctcacagatctcctcctctgtctcgtacgtcgatccgaac 149  
QY 181 tctcagatcttgtgtcagccagccagagcagatggtatctctctctcctcgc 240  
DB 150 tctcagatcttgtgtcagccagccagagcagatggtatctctctcctcgc 209  
QY 241 aactcctcctgt 300  
DB 210 aactcctcctgt 269  
QY 301 aaacatgt 360  
DB 301 aaacatgt 360

DB 270 aaacatgt 329  
QY 361 ctgagcatttgaattcattcattcagcagcctcgcagagcgtctacacccgttgacc 420  
DB 330 ctgagcatttgaattcattcattcagcagcctcgcagagcgtctacacccgttgacc 389  
QY 421 tgggtgggaaatttgggaaatacaccagcttgcggggaattgtgactctgtgt 480  
DB 390 tgggtgggaaatttgggaaatacaccagcttgcggggaattgtgactctgtgt 449  
QY 481 ccttggtgtgttccatgagcgttgccaaagcttgggagcagctcagaaatctacaat 540  
DB 450 ccttggtgtgttccatgagcgttgccaaagcttgggagcagctcagaaatctacaat 509  
QY 541 gtcaagctgtctgtcagacagcagagcagctgtgtgtgtgtgtgtgtgtgtgtgt 600  
DB 510 gtcaagctgtctgtcagacagcagagcagctgtgtgtgtgtgtgtgtgtgtgtgt 569  
QY 601 gtgacgtgagatcccaaggcagggcctatgtgtgaaagcaacttaccatgctgtgc 660  
DB 570 gtgacgtgagatcccaaggcagggcctatgtgtgaaagcaacttaccatgctgtgc 628  
QY 661 ccagctcttgatcacctgtcagacacacacagcagcagaattgtgtctagctgtgtgt 720  
DB 629 ccagctcttgatcacctgtcagacacacacagcagcagaattgtgtctagctgtgtgt 688  
QY 721 aactgctgtcctgaactctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780  
DB 689 aactgctgtcctgaactctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 748  
QY 781 atccttcaacttggatccattacagcagccagccagcagcagcagcagcagcagcagc 840  
DB 749 atccttcaacttggatccattacagcagccagccagcagcagcagcagcagcagcagc 808  
QY 841 gtcctatcgaacacacagctgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900  
DB 809 gtcctatcgaacacacagctgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 868  
QY 901 agagagcaagtatgt 960  
DB 869 agagagcaagtatgt 928  
QY 961 gctctgagggcagcagcctctccttgcgggaataagccttcaacttctgcagacttgc 1020  
DB 929 gctctgagggcagcagcctctccttgcgggaataagccttcaacttctgcagacttgc 988  
QY 1021 gccgacacctgtcctgt 1080  
DB 989 gccgacacctgtcctgt 1048  
QY 1081 caatgtcttaagactgtactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140  
DB 1049 caatgtcttaagactgtactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1108  
QY 1141 acctatgagtaatgt 1200  
DB 1109 acctatgagtaatgt 1168  
QY 1201 gagaacctgtactctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260  
DB 1169 gagaacctgtactctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1228  
QY 1261 gagcttcaactgt 1320  
DB 1229 gagcttcaactgt 1288  
QY 1321 tatcttcaagaaatgt 1380  
DB 1289 tatcttcaagaaatgt 1348  
QY 1381 taatcataaagtcaaaaaaataaaaaaataaaaaa 1414  
DB 1349 taatcataaagtcaaaaaaataaaaaaataaaaaa 1382

RESULT 3  
AA30398  
ID AA30398 standard; DNA; 1203 BP.  
XX  
AC AA30398;  
XX  
DT 14-MAY-1999 (first entry)  
XX  
DE DNA encoding a human secreted protein.  
XX  
KW Secreted protein; cancer; tumour; neurodegenerative disorder;  
KW developmental abnormality; foetal deficiency; blood disorder;  
KW CNS disorder; immune system disease; autoimmune disease; hepatic  
KW renal disease; diabetes; inflammation; allergy; ischemic shock;  
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;  
KW prostate disease; asthma; osteoporosis; arthritis; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO9907891-A1.  
XX  
PD 18-FEB-1999.  
XX  
PF 04-AUG-1998; 98WO-US16235.  
XX  
PR 19-AUG-1997; 97US-0056732.  
PR 05-AUG-1997; 97US-0054798.  
PR 05-AUG-1997; 97US-0054803.  
PR 05-AUG-1997; 97US-0054804.  
PR 05-AUG-1997; 97US-0054806.  
PR 05-AUG-1997; 97US-0054807.  
PR 05-AUG-1997; 97US-0054808.  
PR 05-AUG-1997; 97US-0054809.  
PR 05-AUG-1997; 97US-0055309.  
PR 05-AUG-1997; 97US-0055310.  
PR 05-AUG-1997; 97US-0055312.  
PR 05-AUG-1997; 97US-0055386.  
PR 05-AUG-1997; 97US-0055911.  
PR 18-AUG-1997; 97US-0055970.  
PR 18-AUG-1997; 97US-0055986.  
PR 19-AUG-1997; 97US-0056365.  
PR 19-AUG-1997; 97US-0056366.  
PR 19-AUG-1997; 97US-0056557.  
PR 19-AUG-1997; 97US-0056370.  
PR 19-AUG-1997; 97US-0056371.  
PR 19-AUG-1997; 97US-0056563.  
PR 19-AUG-1997; 97US-0056731.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Brewer LA, Ebner R, Ferlie AM, Greene JM, Janat F, Ni J;  
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;  
XX  
DR WPI: 1999-167452/14.  
XX  
DR P-PSDB: AAY10877.  
XX  
PT New isolated human genes encoding secreted polypeptides - useful for  
XX  
PT diagnosis and treatment of pathological diseases  
XX  
PS Claim 3; Page 265-266; 331pp; English.  
XX  
CC The specification describes secreted proteins and their corresponding  
CC polynucleotides which are useful for preventing, treating or ameliorating  
CC medical conditions, e.g. by protein or gene therapy. Pathological  
CC conditions can also be diagnosed by determining the amount of the  
CC secreted polypeptides in a sample or by determining the presence of  
CC mutations in the polynucleotides. Specific uses are described for each  
CC of the products, based on which tissues they are most highly  
CC expressed in, and include developing products for the diagnosis or  
CC treatment of cancer, tumours, neurodegenerative disorders, developmental  
CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,

CC diseases of the immune system, autoimmune diseases, hepatic and renal  
CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's  
CC and cognitive disorders, schizophrenia, cardiovascular disorders,  
CC prostate diseases, asthma, disorders involving osteoclasts such as  
CC osteoporosis, arthritis or malignancies, diseases of testes, lung or  
CC thymus, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
XX  
SQ Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;  
XX  
Query Match 63.8%; Score 904; DB 20; Length 1203;  
Best Local Similarity 99.8%; Pred. No. 4e-240;  
Matches 926; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
Qy 466 gtgagacttggtctgcttgggtggttccatgagcgtgagcagagactggagcagactc 525  
Db 1 gtgagacttggtctgcttgggtggttccatgagcgtgagcagagactggagcagactc 60  
Qy 526 agaaatctacaatgttcacgtgctcgtcgaacagcaaaaggcgagtagtgccactaca 585  
Db 61 agaaatctacaatgttcacgtgctcgtcgaacag -aaaggcgagtagtgccactaca 119  
Qy 586 ggaagacacatctgtgtgacgtagagatccagggcagggcgctatgtgtgaagcaact 645  
Db 120 ggaagacacatctgtgtgacgtagagatccagggcagggcgctatgtgtgaagcaact 178  
Qy 646 ctaccatgcttggtggccagctttagtgaactgtgtgacacccagcaagaattgttc 705  
Db 179 ctaccatgcttggtggccagctttagtgaactgtgtgacacccagcaagaattgttc 238  
Qy 706 tagctgtctgtatgacatgctggttccctgaactctctctgtgacatgtgtcctaactgtgag 765  
Db 229 tagctgtctgtatgacatgctggttccctgaactctctctgtgacatgtgtcctaactgtgag 298  
Qy 766 cagagataacttaactatcttccatcagctttagatccatcaacagggcccaactgtgagag 825  
Db 299 cagagataacttaactatcttccatcagctttagatccatcaacagggcccaactgtgagag 358  
Qy 826 tttgtctgctggggcccggtgtctatgaaacccagctgtatgtgtgcaacagacagtgtg 885  
Db 359 tttgtctgctggggcccggtgtctatgaaacccagctgtatgtgtgcaacagacagtgtg 418  
Qy 886 gacgccacacatgagaagaagcaagttatgtgccaacagatgtgtgtagacccttgggaa 945  
Db 419 gacgccacacatgagaagaagcaagttatgtgccaacagatgtgtgtagacccttgggaa 478  
Qy 946 cagtgtgtgcccgtgtctgtgaaggggccaggcctctgtccttgccgaatagaactcaact 1005  
Db 479 cagtgtgtgcccgtgtctgtgaaggggccaggcctctgtccttgccgaatagaactcaact 538  
Qy 1006 atcttgagcagtttgcgcgcagacacgtgcctgtgttccacagccagcgcttaactctatg 1065  
Db 539 atcttgagcagtttgcgcgcagacacgtgcctgtgttccacagccagcgcttaactctatg 598  
Qy 1066 gcaatctggtgcacacacatgtcttaagacttgactctgtgtgatttagacctgcccctcc 1125  
Db 599 gcaatctggtgcacacacatgtcttaagacttgactctgtgtgatttagacctgcccctcc 658  
Qy 1126 cacccccaacctgccaactatagctagctatgtctcatgtgacttggagggcagatccaagcac 1185  
Db 659 cacccccaacctgccaactatagctagctatgtctcatgtgacttggagggcagatccaagcac 718  
Qy 1186 agctccctcaacttggaagaaccttgacactctttagatgaacacagatggcgcttggga 1245  
Db 719 agctccctcaacttggaagaaccttgacactctttagatgaacacagatggcgcttggga 778  
Qy 1246 aagaacttcaacttgagacttcaacttgaggtcagactcagtttcagaaaggttgaattt 1305  
Db 779 aagaacttcaacttgagacttcaacttgaggtcagactcagtttcagaaaggttgaattt 838  
Qy 1306 tataatgacatgttattatcattatgaagaaactgaagttcgtcgtgagggcgtgacagactgg 1365  
Db 1306 tataatgacatgttattatcattatgaagaaactgaagttcgtcgtgagggcgtgacagactgg 1365



Db 839 tataatgcatctgtattcatcagaaactgagctctgtagggctgagcagcactg 898  
QY 1366 cattgaaaataataataatcataagtc 1393  
Db 899 cattgaaaataataataatcataagtc 926

RESULT 4  
AAD25458  
ID AAD25458 standard; cDNA; 1214 BP.  
AC AAD25458;  
DT 26-MAR-2002 (first entry)  
DE Xenopus laevis Nrl1 cDNA.  
XX  
XX Frog; Nrl1; cytosolic; neuroprotective; cellular pathway; therapy;  
KM apoptosis; proliferative disorder; degenerative disease; ss.  
XX  
OS Xenopus laevis.  
XX  
XX Key Location/Qualifiers  
FH CDS 180..1046  
FT /\*tag= a  
FT /\*product= "Frog Nrl1 protein"  
XX  
XX WO200187958-A2.  
XX  
XX 22-NOV-2001.  
XX  
XX 15-MAY-2001; 2001MO-US15664.  
XX  
XX 16-MAY-2000; 2000US-204713P.  
XX  
XX (UYJE-) UNIV JEFFERSON THOMAS.  
XX  
XX PI Croce C, Brenner C, Pekarski Y;  
XX  
XX WPI; 2002-082984/11.  
XX  
XX P-PSDB; AAE15792.  
XX  
XX Isolated cDNA encoding human, mouse, frog and yeast Nrl2 proteins.  
PT useful to find molecules that mimic or antagonize Fhl1 interaction for  
PT the treatment of proliferative or degenerative diseases  
XX  
XX Claim 21; Page 60; 61pp; English.  
XX  
XX PS The invention relates to isolated nucleic acids comprising a fully  
CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and  
CC mouse Nrl2 proteins. Nrl and Fhl1 proteins are encoded as fusion  
CC proteins in invertebrates and as separate polypeptides in vertebrates.  
CC Nrl and Fhl1 interact physically and functions in same cellular pathways.  
CC Molecules which bind Nrl2 and mimic or antagonise Fhl1 interaction are  
CC used to treat diseases in which activity of Nrl2 protein is altered in  
CC a mammal. Fhl1 mimics induce apoptosis and are particularly useful to  
CC treat proliferative disorders, whilst Fhl1 antagonists promote cell  
CC proliferation and are particularly useful to treat degenerative disease.  
CC The present sequence is frog Nrl1 cDNA.  
XX  
XX SO Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;

Query Match 25.2%; Score 356.2; DB 24; Length 1214;  
Best Local Similarity 64.5%; Pred. No. 2.3e-88;  
Matches 532; Conservative 0; Mismatches 293; Indels 0; Gaps 0;

QY 245 gccctctgtgctgtgctgacaggaatcagccagcagcagaagaagaactttaaac 304  
Db 197 gccctctgtgctgtgctgacaggaatcagccagcagcagaagaagaactttaaac 256  
QY 305 atgtgcagactggttcgagagctgacagactggtgcctgctgcttcctgctga 364  
||| | |||| | | |||| | | | |||| | | |||| |

Db 257 gtcttcgagctgatacccgagagctgcggggcgtgcgcctgtgcatgtgtttctgcgga 316  
QY 365 ggcatttgacttcattcagcgggacccctgcagagagctaacctgtctgaaccactgg 424  
Db 317 agcctttgactatactcggggcagcatttgagagcctgagttcgtgactctaca 376  
QY 425 tgggaactcttgggaataataccagcttcgacaggaatgtgacttcgtctcctt 484  
Db 377 tggggaacacattcagaggttaacacccaactgcagagagatgtgtgcttcgtccct 436  
QY 485 ggggtgttccatgagctggtgcgaagacttggagcagactcagaataactaattgtca 544  
Db 437 ggggggatttcattcagaaagagcccaacttggacacgacccaacttccatctca 496  
QY 545 cgtgctctgaacagcagaagggcagtagtggccacttaacaggaagacatctgttga 604  
Db 497 cgtgtgtgtgacacacagcagacagtagatcgtgttacccgcaaggctccactgttga 556  
QY 605 cgtagagattccagggcagggcctatgtgtgaagcgaactcaccatctgcctggccag 664  
Db 557 cgtagacttgcagaaatgtagtgcactcagagagcagttccaccctcccgagcaga 616  
QY 665 tctgaagtcacctgtgcagcacaccagcagcagaagattgtctagctgtctatgaat 724  
Db 617 gcttattcgcccatcactcactcactcagcaggaagattggcctgggggtgttagcact 676  
QY 725 gcggttccctgaactctctctgtgcatgtgtcgaagctggagcagagataactctatcc 784  
Db 677 ccgcttcccaaatctctcctgtgtcgtgcgccaacagagcagaactctcacttacc 736  
QY 785 ttcagctttgagatcattcaaggccagcccaacttgggaggtgtgtctgcgggcccgtgc 844  
Db 737 ttcgcttccacccctcactactgtctgcacattggagaggtgtgtctgtagagccgtgc 796  
QY 845 tatcgaaccacagtgctatgtagtgagcagcagcaatgttgaagccacacatgaagaag 904  
Db 797 catagaaacccagtgctagtgatgtgcagcggcagcagacagacacacatgaagaag 856  
QY 905 agcagatataggcacagcagatgtgtgtagaccctcggggagacagtggtgcgcgtctc 964  
Db 857 gacgtcctatgtlcaagctatgtgtgtagaccgttggggctgtgtlcatlgtgccaatgca 916  
QY 965 tgaaggccagagcctctgccttcgccaatgaactcaactctgtgacaggttgcgcg 1024  
Db 917 ggaaggacagaataatgattatgtatgtatgtatgtatgtatgtatgtatgtatgt 976  
QY 1025 acactgcctgtgttccagcacccgagcctgacactctatgtgcaa 1069  
Db 977 ggaatgcggtgtgtgagggcacccgacgactgactgtatgtgaa 1021

RESULT 5  
ABL12225  
ID ABL12225 standard; cDNA; 1495 BP.  
XX  
XX ABL12225;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001MO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX  
XX



QY 253 tggctgtgcccagtgtaacatcgacgcccagacaagacaagaacttaaaactgtgctg 312  
 DB 1200 tttctgtgtggtcagatgctgtcctacacagcgagcgttaacttaactgaagtagatg 1259  
 QY 313 agctgtgtcagagagctgcagactgtggtgctgctgtgcttccctgaagcattg 372  
 DB 1260 agctagtgtatagggccaagtacagaacgctgcagctcttcttctgctgagtgctg 1319  
 QY 373 acttcatgacagggccctgcagagacgctacacgtgtctgaacacactgtggtggaac 432  
 DB 1320 acttctgtggtcagagagcggaccacaactatttgactctccggaggtgtgagcgag 1379  
 QY 433 ttttggaaataacacccagctgtccagggaaatgtgactctgtgctgtctgtgtgt 432  
 DB 1380 taatgtgcagatgcagggaattgtgcgaatgtcacaaaatttggattccctgtgtgcg 1439  
 QY 493 tccatgagcgtgtgcccgaagctgtggagcagactcagaataacttaacttgcaggtgtgc 552  
 DB 1440 tgcacgagcggaacga-----tcaaaaatcttcaacgctcattgtttgc 1484  
 QY 553 tgaacagcaaaagggtcagtagtggcactacagagaagacacatctgtgtgacgtaga 612  
 DB 1485 tcaacagaaaagggtcagtagcagcagtagtatacagaagaagctgcacatgtttgtataga 1544  
 QY 613 ttccaggtcagaggtcctatgtgtgaagcaacctacacatgctgtggccagtcctgagt 672  
 DB 1545 cttaaa---agttctgcctacagcgaatcagatcagatcagtaacggtccgggtactgtctgagc 1601  
 QY 673 caactctcagcacacagcagcagaagattgtgtcagctgtctgtcgtatgacgtgtcc 732  
 DB 1602 gcccaatgtgagcactccagtttgcagatagaggttcaagatttgcagactcgttttg 1661  
 QY 733 ctgaactctctcgtcagcttgcgtcaagcttggagcagagataacttaactcttcagct 792  
 DB 1662 ctgagcgcggcgtgtcgtctcagagaagctgtgtgccaattgtttaacataccatccgcat 1721  
 QY 793 ttggtatccattacagccagccagcactgtggaggtgtgtgtcgtgggcccgtgcatacga 852  
 DB 1722 tcaacatcgaacacggttaaggcgcacttgggaaatctctcgtcgggccaagggcatagaga 1781  
 QY 853 ccagatgcatgtagtggtgcagcagcagcagcagctgtgagccccccttggaagaagcagaatt 912  
 DB 1782 ctcaatgcttctgtgtcgtcgtgcgtcagataggttgcgcaaacgagaagcagagaagt 1841  
 QY 913 atggtcagacagctgtgtgtagaaccccttgggaaacagtggtgc---ccgtgtcttgaag 969  
 DB 1842 ggggccaacagcatgctgttagcccttgggaaacgtaactgctgtgactgcagcgagcag 1901  
 QY 970 ggtcagagctctgtccttgcgcgaatagacctacaactatctgcgaacgttgcgcgaac 1029  
 DB 1902 agttatataagcagcggccgaaggtggaaccttccgtgtccttcaactctcttaccaga 1961  
 QY 1030 tgcctgtgttcagcagccgagcctgacactatgac 1067  
 DB 1962 tgcctcgtctggaacatcgttcgaagaacatctaacgcc 1999  
 RESULT 7  
 ABL20222  
 ID ABL20222 standard; DNA; 5692 BP.  
 XX  
 AC ABL20222;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical; gene; ds.  
 OS Drosophila melanogaster.  
 XX

PN MO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PMD, Myers EM;  
 XX  
 DR WPI; 2001-658660/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 12139; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;  
 XX  
 Query Match 12.9%; Score 182.8; DB 23; Length 5692;  
 Best Local Similarity 54.4%; Pred. No. 5.3e-40;  
 Matches 445; Conservative 0; Mismatches 352; Indels 21; Gaps 3;  
 QY 253 tggctgtgcccagtgtaacatcgacgcccagacaagacaagaacttaaaactgtgctg 312  
 DB 2859 tttctgtgtggtcagatgctgtcctacacagcgagcgttaacttaactgaagtagatg 2918  
 QY 313 agctgtgtcagagagctgcagactgtggtcgtcgtcttccctgaagcattg 372  
 DB 2919 agctagtgtatagggccaagtacagaacgctgcagctcttcttctcgtgagtgctg 2978  
 QY 373 acttcatgacaggggaccctgcagagacgctacacgtgtctgaacacactgtggtggaac 432  
 DB 2979 acttctgtggcgaagcggaccgaacccaactatgtgctctccgaggttgtgagcgag 3038  
 QY 433 ttttggaaataacacccagctgtcagggaaatgtgactcgtcgtcttccctgaagcattg 492  
 DB 3039 taatgtgcagtagtaccgggaattgtgcgaatgcaacaatttggattccctgtgtgcg 3098  
 QY 493 tccatgagcgtgtgcccgaagctgtggagcagactcagaataacttaacatgtcaactgtgc 552  
 DB 3099 tgcacgagcggaacga-----tcaaaaatcttcaacgctcattgtttgc 3143  
 QY 553 tgaacagcaaaagggtcagtagtggcactacagagaagacaactctgtgtgagctaga 612  
 DB 3144 tcaacagaaaagggtcagtagcagctacacgaagaagctgcacatgtttgtataga 3203  
 QY 613 ttccaggtcagaggtcctatgttgaagaacacttaccatgtcgtggccagcttgaat 672  
 DB 3204 cttaag---agttcgtcctacgacgaatcagatagatgaagtcgacgggaatcgtctgagc 3260  
 QY 673 caactgtcagcacacagcagcagaagattgtgtcagctgtcgtctatgacatgcgtgtcc 732  
 DB 3261 gcccaatgtgagcactccagttgtgcgaatagagcttcaagatttgcagactcgttttg 3320  
 QY 733 ctgaactctctcgtcagcttggcctcaagcttggagcagagataacttaacttaacttgc 732

```
Db 3321 cgaagccggcgggtgctcctcaggaagctggtgccatttgttaacataccatccgcat 3380
Oy 793 ttggatccattacagcccccagcactgggaagtggtgtgctggggccgtgtcatcga 852
Db 3381 tcaataagcaaccggttaagcgccacttggaaatcctctcgcgccacagcatagaga 3440
Oy 853 cccagtgctatgtagtgcagcagcaagtgtagccacacatgagaagaagcaagtt 912
Db 3441 cccaatgcttggtgtcgctgcgcagatagttgtgcacacacagcagaagcagatt 3500
Oy 913 atggccacacagctatgtgtgtagaacctctggggaacagtgtagc---ccgctgtctgag 969
Db 3501 ggggcccacacagcatgctgtagcccttggggaacgtagctgtgactgcagcgagcag 3560
Oy 970 ggcacagcctctgcttgcctgcagatagacctcaactatgcgcagacgttgcgcgcagcc 1029
Db 3561 agcttgatataagcagcgccgaaggtggaaccttccgttcgaatcctctgtatcagaca 3620
Oy 1030 tgcctgtgttcacagcagcgacgctgacctctatgac 1067
Db 3621 tgcctgtgttcgaacatcgttcgaacacatctacgccc 3658
```

## RESULT 8

AAD25457 standard; cDNA; 1292 BP.

AAD25457;

26-MAR-2002 (first entry)

Mouse Nlt2 cDNA.

Mouse; Nlt2; cytosolic; neuroprotective; cellular pathway; therapy; apoptosis; proliferative disorder; degenerative disease; ss.

Mus sp.

Key Location/Qualifiers

FT CDS 46..876

FT /tag= a

FT /product= "Mouse Nlt2 protein"

PN WO200187958-A2.

PD 22-NOV-2001.

XX 15-MAY-2001; 2001WO-US15664.

XX 16-MAY-2000; 2000US-204713P.

XX (UVE-) UNIV JEFFERSON THOMAS.

XX PI Croce C, Brenner C, Pekarski Y;

XX WPI; 2002-082984/11.

XX P-PSDB; AAE15791.

XX Claim 9; Page 59-60; 61pp; English.

CC The invention relates to isolated nucleic acids comprising a fully  
CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and  
CC mouse Nlt2 proteins. Nlt and Phit proteins are encoded as fusion  
CC proteins in invertebrates and as separate polypeptides in vertebrates.  
CC Nlt and Phit interact physically and functions in same cellular pathways.  
CC Molecules which bind Nlt2 and mimic or antagonise Phit interaction are  
CC used to treat diseases in which activity of Nlt2 protein is altered in  
CC a mammal. Phit mimics induce apoptosis and are particularly useful to

CC treat proliferative disorders, whilst Phit antagonists promote cell  
CC proliferation and are particularly useful to treat degenerative disease.  
CC The present sequence is mouse Nlt2 cDNA.

XX Sequence 1292 BP; 354 A; 292 C; 289 G; 357 T; 0 other;

Query Match 8.0%; Score 113.8; DB 24; Length 1292;  
Best Local Similarity 51.7%; Pred. NO. 3.4e-21;  
Matches 284; Conservative 0; Mismatches 262; Indels 3; Gaps 1;

```
Oy 516 ggcagactcgaataatctacaattgtcagctgctgtgtaaacagaaggcagtagtg 575
Db 307 gagatgtcgggaaactgtataatccctgtctgtgttggcctgatatggaagttacgg 366
Oy 576 gccacttaaggaagaacacatctgtgtgacgttagaagattccagggcaggtcattgt 635
Db 367 gtaaaagacaaggaagatccatctgtttgacatgtatgttccctgggaaattcgtttcaa 426
Oy 636 gaagacaactcaccatgctgtggccagctctgtagtcacgtgtcagcacaccagcagggc 695
Db 427 gatctaaacattgagccctgtgtatagttctccacatttgat---agccttactgc 483
Oy 696 aagattgtctagctgctgtctatgacatgaggttccctgaaactctctctgtgcatgtgt 755
Db 484 aagatggcctgtgcatctgtcatgatatgacgttgcggagctgtgacacaaatctatgca 543
Oy 756 caagcctggagcagagatactactcctcctcagcttggatggatccattacagcagccagc 815
Db 544 caaagagctccagcagcttgtgtatccttgaggttcaatctacacagcagcagccagcc 603
Oy 816 cactggagagtggtgtcgtcgccgctgtgctatcgaacccagtgctatgtatgtgcaaga 875
Db 604 cactggagagctgtctcagcagagccggcgtgtgtatataacagtgtagtgtgctacagcc 663
Oy 876 gacagtgtagcagccacatgagaagagcagatgtagccacagcagctgtgtgtagac 935
Db 664 tctcctgctcgagtagacaagcctctgtatgtgctgtgggacacagcactgtgtgat 723
Oy 936 ccttggggaacagtggtggtccgctgtgctctgaggggcccagcgtctgtgctccgaata 995
Db 724 ccttgggggcaagctcctaaccaagcgtgcagcaggaagaacatctgttaactcagaata 783
Oy 996 gacctaacatctgcgaagtgtagcgcgacacactgtgctgtgttccagcaccgagcct 1055
Db 784 gacctgaagaagctgtgtgaattcgcgcagcaatcccatcttaaacagagaagagca 843
Oy 1056 gacctctat 1064
Db 844 gacctctat 852
```

## RESULT 9

AAF84214 standard; cDNA; 1001 BP.

AAF84214;

12-JUN-2001 (first entry)

Human hNlt3-ase coding sequence.

Human; hNlt3-ase; Cushings adrenal gland tumour;

Human; hNlt3-ase; Cushings adrenal gland tumour;

Human; hNlt3-ase; Cushings adrenal gland tumour;

Human; hNlt3-ase; Cushings adrenal gland tumour;

Human; hNlt3-ase; Cushings adrenal gland tumour;

Human; hNlt3-ase; Cushings adrenal gland tumour;

Human; hNlt3-ase; Cushings adrenal gland tumour;

Human; hNlt3-ase; Cushings adrenal gland tumour;





CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB01840-AB16175) and the encoded proteins  
CC (AB57737-AB572022).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 898 BP; 243 A; 236 C; 231 G; 188 T; 0 other;

Query Match	7.1%	Score 101.2;	DB 23;	Length 898;
Best Local Similarity	50.6%	Pred. No. 8.8e-18;		
Matches 272; Conservative	0;	Mismatches 263;	Indels 3;	Gaps 1;

OY	551	atctacaattgtcacgctgctgtcgtaaacagcaagagcgagtagtggccacttcaggaag	590
Db	344	atctcaacaacactgcacgctgtttgttcacaccactgtgtaaccttggccaagatcgcgaag	403
OY	551	acaactctgtgtgacgtagagattccacgggcaggggcctctgtgtgaagcaacttacc	650
Db	404	atgcactctcttggacataagatgtctaaagg---tggcattcgtctcaagsgatctggaacg	460
OY	651	atgctctgggcccagctcttgtagtaacctgtcaagcaccacagcaggaagtgtctagct	710
Db	461	ctgttcgcaggaacatgatattccatcatccaacgttagatggccaagaagtgtgcaltcgcg	520
OY	711	gtctgcctatgaacatgcggttccctctgaactctctctgtgcattggtctcaagcttgagcagag	770

Oy	771	atactacacatcccttcaggttttggatcatcatcaaggccagccacgtggagagtg	830
Db	561	atgcatctatcccgctgtgcatcacaacagacacacgtgcccacgtgagagctatg	640
Oy	831	ctgggagcccgctgctatccgaacccacgtgtctatgtagtgcagcagacagatgtgtgaagtc	890
Db	641	cagcgatcccgctgccaatgacaaccaactcttgtgtgcacaacatcaacagcccgtagac	700
Oy	891	caccatgagagagagaagtattatgcccacagatgtgtgtgaacccctggggaacagtg	950
Db	701	acaagccgcagagatgtagagcctatgcccattccatgtagtgtgaatccatgagcccaagtg	760
Oy	951	gtggcccgctgctctgcagggggcagagccctctgcccctgcaggaatgagccatcaatctg	1010
Db	761	cagcagagtgctccagtgtagagggcaggaatcgtgtgtgccgatatgattctcccgagtg	820
Oy	1011	cgacagtgccgcagacaacacgtgctgtgttcacagacccgcagagcctgacactatgga	1068
Db	821	gagcagtgctgtcagcagatctcccgctcttctgggacaagaagcgtctagatctgaccca	878

RESULT	13
ABL13786	
ID	ABL13786 standard; cDNA; 2953 BP.
XX	
AC	ABL13786;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 35840.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
FN	W0200171042-A2.
XX	
PD	27-SEP-2001.
XX	

23-MAR-2001; 2001WO-US09231.  
23-MAR-2000; 2000US-191637P.  
11-JUL-2000; 2000US-0614150.  
(PEKE ) PE CORP NY.  
Venter JC, Adams M, Li PWD, Myers EW;  
WPI; 2001-656860/75.  
P-PSDB; ABB69683.  
New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  
Claim 1; SEQ ID NO 35840; 21pp + Sequence Listing; English.  
The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB161517-AB130511), expressed DNA sequences (AB101840-AB161517) and the encoded proteins (ABB57737-ABB72072).  
The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
Sequence 2953 BP; 692 A; 688 C; 776 G; 797 T; 0 other;

Query Match	7.1%;	Score 101.2;	DB 23;	Length 2953;
Best Local Similarity	50.6%;	Pred. No. 1.6e-17;		
Matches 272;	Conservative 0;	Mismatches 263;	Indels 3;	Gaps 1;

OY	531	atctacaattgtc	aagctgtcgtc	ctgcaacgca	aagggcga	gtacgtatgtg	ccacttacg	gaag	59
Db	1399	atcttacaacac	cttgcaaggtt	gtgtccaccac	ctgtgtacc	cttgtgtgtg	ccaaagcat	gtcgaa	14
OY	591	acacatctgtg	tagcaatg	atccaaaggtg	caagggcctaa	gtgtgtgaag	caactctaac	65	
Db	1459	atgcattctctt	gtacataatg	tgtccaaggg	-----	tggcattgc	cttcaagga	gtcggaaacg	15
OY	651	atgctctgag	cccaagctct	gtgaagtcac	ctgtcaagca	ccagcagga	aatgtgtac	71	
Db	1516	ctgtccgcaag	gcaaaagatt	caaccatca	ccaagctaga	tgtgcaca	aaagattgtg	ctcggc	15
OY	711	gtctgcata	gtgcaatg	ctgggttcc	ctgtcaactc	ctctctgtga	tgtgtc	caagctgtga	77
Db	1576	atctgtctac	gcgataatc	gcgattcga	ggagatgtg	gcaggtcct	ctatcga	agcaggtc	16
OY	771	atacttacct	atccttcac	gctttgtga	atccatctaac	agcccaag	cccaactgtg	gaagtgtg	83
Db	1636	atgatactcat	ctacccgtc	gtcatcca	acatactga	cccaactgtg	ctccactgtg	caactgtg	16
OY	831	ctgctggg	cccggtgtc	atctgaa	cccaagctgt	atgtaggtg	gaagcag	ctatgtgtg	89
Db	1696	cagctga	atcccggtg	ccaatgtaca	ccaactctt	gtgtgtca	caacatca	cccaagcccg	17
OY	891	caaccctga	agaagagca	agtattgtat	gtccacacga	ctgtgtgtga	gaacccctgtg	gaacag	95
Db	1756	acaagcgc	gcaggtat	gtatgtac	gtgcctat	gtgcattccat	ctgtgtgtga	atccatgtg	18
OY	951	gtgtgccgc	gtctctc	gtgaaggtgc	caaggtcct	ctgtgcctt	gcgaatga	actcaactat	10
Db	1816	cagcaga	agtgtcc	caagtgtga	aggtcga	gaaatcgt	gtgtgtgcga	atataatgtat	18
OY	1011	cgaacag	ctgtgcgc	gcgaacac	ctgtcctgt	gtctccacga	cccgcaag	ctgtac	1068
Db	1876	gaagcag	ctgtgcac	caatctcc	ctcttctgtg	gcaaaac	ctctcaat	ctctatcc	1933





PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249220.  
PR 17-NOV-2000; 2000US-0249221.  
PR 17-NOV-2000; 2000US-0249222.  
PR 17-NOV-2000; 2000US-0249223.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249226.  
PR 17-NOV-2000; 2000US-0249227.  
PR 17-NOV-2000; 2000US-0249228.  
PR 17-NOV-2000; 2000US-0249229.  
PR 17-NOV-2000; 2000US-0249230.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250161.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251859.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI: 2001-483426/52.  
P-PSDB: AAM88455.  
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and  
metastasis -  
PS Claim 1; SEQ ID NO 6296; 3071bp + Sequence Listing; English.  
AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic  
activity, and can be used in gene therapy and vaccine production. (I)  
proteins and polynucleotides may be used in the prevention, diagnosis and  
treatment of diseases associated with inappropriate (I) expression. For  
example, they may be used to treat disorders associated with decreased  
expression by rectifying mutations or deletions in a patient's genome  
that affect the activity of (I) by expressing inactive proteins or to  
supplement the patient's own production of (I). Additionally, (I)  
polynucleotides may be used to produce the secreted (I), by inserting  
the nucleic acids into a host cell and culturing the cell to express the  
protein. (I) proteins and polynucleotides may be used to prevent,  
diagnose and treat immune/haematopoietic-related diseases, especially  
cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
to AAK87694 represent human immune/haematopoietic antigen genomic  
sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 539 BP; 129 A; 117 C; 137 G; 152 T; 4 other;

Query Match 6.7%; Score 95.2; DB 22; Length 539;  
Best Local Similarity 95.9%; Pred. No. 3.2e-16;  
Matches 94; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 109 ggcgtggtcattacacgcgcctccacacattctctccctctgtgtctgtgactcc 168  
|||||

Db 371 ggctgggcttcacaccaggcctccacagattccctgcccctctctgtgckggnctcc 430  
Qy 169 ggatacctcaactctcaactcttctgtgctcagcccaag 206  
Db 431 ggatacctcaactctcaactcttctgtgctcagcccrpg 468

Search completed: April 28, 2002, 04:52:22  
Job time: 3740 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2002, 01:30:45 ; Search time 1522.25 seconds  
(without alignments)  
12554.895 Million cell updates/sec

Title: US-09-357-675C-1  
1416  
Sequence: 1 gccacacgcgcgcgcctnt.....aaaaaaaaaaaaaaaaaa 1416

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	939	66.3	992	9	AL520767
C 2	927.6	65.5	960	9	AL522373
C 3	888.6	62.8	1374	11	AK002269 Mus muscu
4	886.6	62.6	2993	11	AK004988 Mus muscu
5	845	59.7	890	9	AL520768
6	844	59.6	844	9	AL529152
C 7	795.8	56.2	853	9	AL562604
C 8	772.4	54.5	830	9	AL580840
C 9	759.6	53.6	916	9	AL563355
C 10	758	53.5	846	10	AL563355
11	758	53.5	861	9	AL559163
12	748.6	52.9	928	10	AL559163
13	724.4	51.2	793	10	AL559163
14	711	50.2	790	10	AL559163
C 15	704.4	49.7	787	9	AL559163
16	699.2	49.4	764	10	AL559163
17	686.8	48.5	746	10	AL559163

18	678.4	47.9	813	10	BT755493
C 19	666.2	47.0	736	9	AL668782
20	665.2	47.0	946	10	BT757823
21	654.6	46.2	791	10	BT776560
22	643.4	45.4	781	10	BT532265
C 23	638	45.1	710	9	AT1797380
C 24	625	44.7	625	10	BT111300
C 25	619.4	43.7	828	10	BE895409
C 26	616.4	43.5	690	9	AA909728
27	609.2	43.0	875	9	AL347361
28	603.8	42.6	900	10	BF968272
29	598.2	42.2	778	10	BF968495
C 30	589.2	41.6	635	10	BF448631
C 31	584	41.2	586	9	AT1797259
C 32	583.6	41.2	900	10	BT432389
C 33	579.8	40.9	612	9	AT1831417
C 34	577.2	40.8	657	10	BT1916460
35	572.8	40.5	827	10	BF9696074
36	572	40.4	579	10	BT111712
37	557.4	39.4	576	9	AW956706
C 38	557	39.3	559	10	BT411822
C 39	554.2	39.1	579	9	AT159188
C 40	552.8	39.0	560	9	AW182514
C 41	548.2	38.7	869	10	BT963013
C 42	548	38.7	550	10	BT411736
C 43	547.4	38.7	550	9	AT1017543
C 44	547.4	38.5	832	9	BT647584
45	535.6	37.8	672	10	BE392424

#### ALIGNMENTS

RESULT 1	AL520767	992 bp	mrna	linear	EST 13-FEB-2001
LOCUS	AL520767	LTI_NFL004_NBC2	Homo sapiens	cdna	clone CS0DB002YD05 3
DEFINITION	AL520767	LTI_NFL004_NBC2	Homo sapiens	cdna	clone CS0DB002YD05 3
ACCESSION	AL520767	prime, mRNA sequence.			
VERSION	AL520767.1	GI:12784260			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.				
JOURNAL	Full-length cDNA libraries and normalization				
COMMENT	Unpublished (2001)				
CONTACT	Contact: Genoscope				
Genoscope	Genoscope - Centre National de Sequencage				
BP 191	91006 EVRY cedex - France				
Email:	segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
Location/Qualifiers					
1. 992					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone_lib="CS0DB002YD05"					
/sex="male"					
/tissue_type="neuroblastoma cells"					
/lab_host="DH10B"					
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cdna					
was primed with a NotI-oligo(dT) primer. Five prime end					
cloned, double-stranded cdna was digested with Not I and					
enriched into the Not I and Eco RV sites of the pCMVSPORT 6					
vector. Library was normalized. Library was constructed					
by Life Technologies. Contact: Feng liang life					
Technologies, a division of Invitrogen 9800 Medical Center					
Drive Rockville, Maryland 20850, USA fax : (1) 301 610					
8371 Email : filiang@life.com url :					
http://fulllength.invitrogen.com"					

BASE COUNT

229 a 267 c 268 g 222 t 6 others



OY	758	agcttgagcagagataacttacctactccttcagcttttgatccattcaaggcccaagccca	817
Db	603	AGCTGAGACAGAGATACTTACTTACTTCCCTTCACGCTTTTGATTCATTACAGGCCAGCCCA	544
OY	818	ctgggaagtgatgtgcctggcgcccgctgcatacgaaacccagtgctaigtatgtgcagcagc	877
Db	543	CTGGGAGGCTGTTGCTGGCGGCCCGTGCATCATCAAAACCCAGTCTATTGTAGTGGCAGCAGC	484
OY	878	acagtgtygacgcgcacccatlgaaagaagacaagtatagcccaagcagatgtygtlagacc	937
Db	483	ACAGTGTGAGACGCCACCATGAGAGACAGCAAGATTATGSCCAGACATGNTGTAGACC	424
OY	938	ctggggaacagtgatggtggcccgctgtcttcgagggcgcaagcctctgccttgcggcaatga	997
Db	423	CTGGGGAACAGTGGTGGCCCGCTGCTGTCGAGGGGAGAGCCCTTCCTCTTGGCCCGAATAGA	364
OY	998	cctcaactatctcgagcaagttgcgcgcgaacactgctctgtatcacaagcccgagacctga	1057
Db	363	CCTCAACTATCTGCGCACAGTTGCGCCGACACTGCTGCTGTTCGACGACCCGAGGCTGA	304
OY	1058	cctctaigtgcaactctggtgtcaaccactgtcttaagaacttgacttctgtgagttlagacct	1117
Db	303	CCTTATGTGGAATCTGGGTCACCCACTGCTTAAGACTTGACTTGTGAGTTTAGACT	244
OY	1118	gcccctcccaaccccaaccctgcacctaagagctagtgctcatgtgtaacttggagcgagat	1177
Db	243	GCCCCCTCCACCCCCACCCCTGCACATTAAGACTGATGCTCATGTGACTTGGAGCGAGGAT	184
OY	1178	ccaggcacaagctcccccactactttggagaaacttgaactccttgatgagaaacaagatggct	1237
Db	183	CCAGGCACAGCTCCCTCTMCCTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGCT	124
OY	1238	gcttgggaagaagaacttcaccactgagcttcacactgaagtcagacgtcagttccagaagcg	1297
Db	123	GCTTGGGAAGMACTTTCACCTGAGCTTCACCTGAGGTGACAGCTTCAGAAAGG	64
OY	1298	tggaaattatataagtcattgtttatttcattgaagaaactgaagttctgtcagagctgagc	1357
Db	63	TGGAAATTTATATAGTCATTGTATTTCATGAGAACTGAAGTCTGCTGAGAGGTTGAGC	4
OY	1358	ag 1359	
Db	3	AB 2	
RESULT	3		
LOCUS	AK002269	1374 bp	mrna
DEFINITION	Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610007B07:nitrlase 1, full insert sequence.		
ACCESSION	AK002269		
VERSION	AK002269.1	GI:12832130	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus. musculus (strain: C57BL/6J) adult male kidney cDNA to mrna, clone_1.lib:RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2 (sites)		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, H., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		

PUBMED AUTHORS	11042195 3 (sites) Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,Y., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamanoto,R., Matsunoto,H., Sakeguchi,S., Ikegami,T., Kashiwagi,K., Fujikawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., -Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL MEDLINE PUBLISHED	20530913 11076861 4 (sites)
REFERENCE AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE JOURNAL REFERENCE AUTHORS	Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5 (bases 1 to 1374) Adachi,T., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoaka,T., Horii,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numataki,R., Ono,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Salto,H., Satto,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schraml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shinkai,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamakoshi,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE JOURNAL	Direct Submission Submitted (10-JUN-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details.
FEATURES SOURCE	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GGAAGAGAAGCGCCGCCAACCTCAATTTTATTATTAATTAACCCCCCCC 3'] cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAAGAGAGAAGCATCAAGAGCTCAATTAATTAATTAACCCCCCCC 3']. cDNA was cloned with XhoI and SstI. Cloning sites, 5' end; SstI; 3' end; XhoI. Host: SOLR.
FEATURES SOURCE	Location/Qualifiers 1. .1374 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="MGI:MGI:1891931" /db_xref="taxon:10090" /clone="0610007B07" /sex="male" /tissue_type="kidney" /clone_id="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult"
gene	1. .1374 /gene="Nict1"
misc_feature	1. .1374 /gene="Nict1" /note="Data source:MGI, source key:MGI:1350916,"



MEDLINE	20530913
PUBMED	110766861
REFERENCE	4 (sites)
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCES	5 (bases 1 to 2993)
AUTHORS	Adachi,J., Alizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bonio,H., Brownstein,M., Bull,C., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanganaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Horii,F., Humo,D., Imocant,K., Ishii,Y., Itoh,M., Izawa,M., Katsukawa,T., Kato,H., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kuhihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schirral,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunushi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGAGAGAAGCGCCGCCACATCGACTGTCTTTTCTTTTIVN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5'] GAGAGAGAGAGATCCAGACGCTCAATTATTAATTAACCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES	Location/Qualifiers 1..2993
SOURCE	/organism="Mus musculus" /strain="C57BL/6J" /db_xref="WGD:MGI:1895066" /db_xref="taxon:10090" /clone="1300013F05" /sex="male" /tissue_type="liver" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 169..1041 /gene="Niti1" 169..1041 /gene="Niti1" /note="data source:WGD, source key:MGI:1350916, evidence:ISS nitrilase 1 putative" /codon_start=1 /protein_id="BAB23723.1" /db_xref="GI:12836591" /translation="MSSSTSWELPLVAVCYTSPPNKQENFKTCALVELQEARLIGACITAFLEADPEFIARNPAETLLLSPEPLNDLGQYSOLARCGIMLSLGGHNEGDMOEMOQKINVCYHLMSKGSVASYSKYTHCIDPVEISGGQPMRSNTYTKGGTELPVKTPDAGVGILGVDMRPPELLSIKAOGAELITLVPSFGSVTGPAHNEVLIRARAIESOCYVTAAACGRHHETRASYSGHSNVDPKWGTVAARCSEGFCLARIIDLHFLOMRHLTFVPQRRRPDYGLGHPPLS"
CDS	polyA_signal 2973..2978

[illegible]

DB	391	GGGCGACAGTGGGCGCCGCTGCTCGGAGGAGCAGGACCCCTGCTTGGTCGCAATTGACCT	950
Qy	1001	caacatctctggacacgtttgcccgcgaacactgcctctgttctcagaacgcgaagcctgacct	1060
Db	951	CCACTTTCTTACAAACAGATAGCCGCAACACCTGCTGTGTTCAGCACCCGACACCTGACCT	1010
Qy	1061	ctatggaactcgtggatcaaccactgtcttaagaactctg--actctctgtatgatttagacctg	1118
Db	1011	CTATGGCAGCTCTGGGTGATCCACTCTCTTAAAGCTTTTAACCTTCTGTTGGTTAAACACACA	1070
Qy	1119	cccccccaacccccaccctctgcacctaagactagctagtctgtcattgtgaactctggagagcagatc	1178
Db	1071	GACACAGC-----CAACTGGGAGGAGTACTGTACTGTGACTGTAAGCAGAGACA	1118
Qy	1179	cagagcagcgtccctcactctgtggagaaccttga-cctccttgatgtgaacacagatggagct	1237
Db	1119	CAGAGCAGGTTCCCTTTCACCTTGCAGAACCTTTAACTCTCTTAAATGGAACACAGGCTTAAT	1178
Qy	1238	gcttggagaagaacttcacactgcagctgcacactgaagtcgaagtcgaagctcagattccagaagc	1297
Db	1179	CAGTCCATGGGAA---AGAGAGACACTTTTGACCTTACCTCAGATTTCAGATTTCACAAAAAG	1235
Qy	1298	tggaaattatataagctatctgttatttcttcattgaaacctgaagttctgctgaagcgctgaagc	1357
Db	1236	TGGATTTTATATAGTCACTGTTTATTTTCATGAAACGTAAGTATATGCTGAGAGCGCTGAGC	1295
Qy	1358	agcactgagcattg----aaaataataataatcataaatc	1393
Db	1236	AGCACTGGCAATTGAAAAAATAATATATATCATATATGTC	1335
RESULT	5		
LOCUS	AL520768	890 bp	mRNA
DEFINITION	AL520768 LTI_NFL004_NBC2 Homo sapiens CDNA clone CS0DB002YD05 5	prime, mRNA sequence.	
ACCESSION	AL520768		
VERSION	AL520768.1	GI:12784261	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eulazia; Primates; Carnivora; Cetartiodactyla; Hominoidea; Homo.		
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.		
TITLE	Full-length CDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.		
FEATURES	Location/Qualifiers		
source	1..890		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CS0DB002YD05"		
	/clone_lib="LTI_NFL004_NBC2"		
	/sex="male"		
	/tissue_type="neuroblastoma cells"		
	/lab_host="DH10B"		
	/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	191 a	251 c	241 g
ORIGIN			206 t
			1 others

Query Match		59.7%	Score	845	DB	9	Length	890
Best Local Similarity		96.4%	Pred.	No. 1	le-132			
Matches		886	Conservative	1	Mismatches	1	Indels	31
								Gaps
								1
Qy	20	tctgtctccagaccgccccccgcgagtcgagacccttgcaatgglttlytgctatacttcacg	79					
Db	3	tctgtctccagaccgccccccgcgagtcgagacccttgcaatgglttlytgctatacttcacg	61					
Qy	80	taggaactactccctatcccgctcggcgctgtgtgtcttaataccagctccctcaag	139					
Db	62	-----	91					
Qy	140	attctgtccctctctgttctctgtgactccggaatacctaactctagtaacttctgtccca	199					
Db	92	attctgtccctctctgttctctgtgactccggaatacctaactctagtaacttctgtccca	151					
Qy	200	gcccaagcccaagaccatggtatctctctctctctgtggaactgtccctgtgtgtc	259					
Db	152	gcccaagcccaagaccatggtatctctctctctctgtggaactgtccctgtgtgtc	211					
Qy	260	gtgtccaggttaaatcgaagccgaagaagaagaactttaaaatgtgtctgagctgt	319					
Db	212	gtgtccaggttaaatcgaagccgaagaagaagaactttaaaatgtgtctgagctgt	271					
Qy	320	tcaagaagctgtccgaactgtgtgtcctgtgacttctctctgtccgtgtgagatctgactat	379					
Db	272	tcaagaagctgtccgaactgtgtgtcctgtgacttctctctgtccgtgtgagatctgactat	331					
Qy	380	tgacagggagacccttgagagacgtctaacctgtctgaaccaactgggttgggaactttga	439					
Db	332	tgacagggagacccttgagagacgtctaacctgtctgaaccaactgggttgggaactttga	391					
Qy	440	agaataaccaccaagcttgacagggaaatgtgagctcgtgtgttccttgggttggtaaga	499					
Db	392	agaataaccaccaagcttgacagggaaatgtgagctcgtgtgttccttgggttggtaaga	451					
Qy	500	gcgttgccaagaacttgagagcagactccagaaaatctacaattgtcagctgtcgtgaacag	559					
Db	452	gcgttgccaagaacttgagagcagactccagaaaatctacaattgtcagctgtcgtgaacag	511					
Qy	560	caaaagggaagtagtgtgccaacttacagaagaacacactgtgtgtagcgttagaatccagg	619					
Db	512	caaaagggaagtagtgtgccaacttacagaagaacacactgtgtgtagcgttagaatccagg	571					
Qy	620	gcaagggactatgtgtgaagcaactctacatgtgctgtggcccaacttlttgatccactgt	679					
Db	572	gcaagggactatgtgtgaagcaactctacatgtgctgtggcccaacttlttgatccactgt	631					
Qy	680	cagcacacacagcagcaaatgtgtctagctgtctgtcatgcatgctgtccctgaact	739					
Db	632	cagcacacacagcagcaaatgtgtctagctgtctgtcatgcatgctgtccctgaact	691					
Qy	740	ctctctgtgcatgtgtcctaagcttgagacagagatacttaacctatccctcgaactlttgatc	799					
Db	692	ctctctgtgcatgtgtcctaagcttgagacagagatacttaacctatccctcgaactlttgatc	751					
Qy	800	catcacaagcccaagcccaactgtggaagtgttctgtgtggcccgctgtatccgaaccacaagt	859					
Db	752	catcacaagcccaagcccaactgtggaagtgttctgtgtggcccgctgtatccgaaccacaagt	811					
Qy	860	ctatctagtggtgagcagacagctgtgagacgcacacatagaagaagtgtaagtltatgagca	919					
Db	812	ctatctagtggtgagcagacagctgtgagacgcacacatagaagaagtgtaagtltatgagca	871					
Qy	920	cagcatgtgtgtagaacctc	938					
Db	872	cagcatgtgtgtagaacctc	890					
RESULT 6								
AL529152 844 bp mRNA linear EST 13-FEB-2001								
LOCUS								



DEFINITION	AL529152.LT1.NFL001.NBC4 Homo sapiens CDNA clone CS0DD004YM06 5 prime, mRNA sequence.
ACCESSION	AL529152
VERSION	AL529152.1 GI:12792645
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Li, W.-B., Gruber, C., Jesssee, J. and Polayes, D.
TITLE	Full-length CDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: <a href="mailto:segre@genoscope.cns.fr">segre@genoscope.cns.fr</a> , Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> .
FEATURES	Location/Qualifiers
SOURCE	1..844
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="CS0DD004YM06"
	/clone_id="LT1_NFL001_NBC4"
	/sex="male"
	/tissue_type="neuroblastoma cells"
	/lab_host="DH10B"
	/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : <a href="mailto:liang@lifetechn.com">liang@lifetechn.com</a> URL : <a href="http://fulllength.invitrogen.com">http://fulllength.invitrogen.com</a> "
BASE COUNT	182 a 238 c 235 g 189 t
ORIGIN	
Query Match	59.6%; Score 844; DB 9; Length 844;
Best Local Similarity	100.0%; Pred. No. 1.7e-132;
Matches 844, Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	300 aaaaatgtctgaagctgtgttcgagaagctgccagaactggctgtcgtcttcctg 359
DB	1 AAAACATGTCGTCGACGCTGTTTCGAGAAGCGCTGCACAGCTGGGTGCTGCTTCCTG 60
QY	360 cctgaagcattgtacttcattgcacggaacctgcagagaagctacacctgtctgaacca 419
DB	61 CCTGAGGCAATTGACTTATTCAGAGGAGACCTCTGCAGAGACGCTACACCTCTGTGAAACCA 120
QY	420 ctgggtggaaacttttgaaagataacacccagctctgcaggaagtgtgactcgtgctg 479
DB	121 CTGGGTGGGAAACTTTTGAAAGAAATACACCAAGCTTGCACAGGAAATGTGACTCGCTG 180
QY	480 tccctgggtgtttcccaagcagctgtgccaagaacttggagcaagcaactcaaaaaactacaat 539
DB	181 TCCCTGGGTGGTTTCCATGACGAGCTGTGCCAAGAGCTGGAGACACACTCAGAAAAATCTACAAAT 240
QY	540 tgtcacgtgtcgtctgaacagcaagaaggagcaagtgtgccaacttacagaagaacacatctg 599
DB	241 TGTCACTGTGAGTCTGTGAAACGCAAAAGGGGCGAGTAGTGCCACACTTACAGGAAGACACATCTG 300
QY	600 tgtgacgtgaagatccagagcgaggggagctatgtgtgaagaagaactctacatctgtctggg 659
DB	301 TGTGACGTGAGGATCCAGGGCGAGGGGCTATGTGTGAAAGCAACTCTACCACTGCTCGG 360
QY	660 cccaatctctgacacctgtcagcacacacagcagcagaagatgtgtctagtctgtcgtat 719
DB	361 CCCAATCTTGAATCACTCTGTACGACACACAGCAGCAAGATGTTGTTGCTTCTGCTGTAAT 420
QY	720 gacatgtgggttccctgaacctctctgtgcaatgtgtcctaagctgtgagcagagatactacc 779

DB	Accession	Version	Keywords	Organism	Source
Db	421	GNATGCGGTTCCCTGGAACCTCTCTGSCATTTGGCTCAAGCTGGAGCAGAGTA	ACTTAC	480	
Qy	780	tatcccttcagccttttga	ttccatc	taagagcccaagccca	ctggaagtgltgctcgagcc
Db	481	TATCCTTCAGCTTTTGGATTCATTCAGAGCCAGCCCACTGGAGAGTGTTGCTGGG	GC	540	
Qy	840	cgctgctacgaaacccag	tgtctatg	taatgagcagagacacag	tgtgtgagccacatag
Db	541	CGTGTATTCGAACCACTGCTATGTATGTATGTGACAGACAGTGTGGACGGCAC	CATGAG	600	
Qy	900	aagagagaagatlatagc	ccacagcatggttg	tagacccttggaagacag	tgtgtgcccgc
Db	601	AAGAGAGCAAGTTATG	CCACAGCATGTTGTTAGACCCCTGGGAAACAGTGTGGCCCCC	GC	660
Qy	960	tgctctgaggggcca	gagcctctg	cccttcgccaatagac	ctcctaactatctcgaaag
Db	661	TGCTGTGAGGGGCCAGGCGCTTGCCTTCCCGAATATGACCTCAACTATCTG	GACAGTTG	720	
Qy	1020	cgcgcagacactg	ctgtgttcc	agcagccagagcctg	acctctatgtgcaatc
Db	721	CGCCACACCTGCTCCCTGTTCAGCAGCCGACAGGCGCTGACCTCTATG	GCAATCTGGGTGAC	780	
Qy	1080	ccacgtctttaa	gacttactctctgtgaggttagaccctgcgcccctccaccccaccc	ctgcgc	1139
Db	781	CCACTGCTTTAAGACTTACTTCTGTGTGATTTAGACCTGCCCCCTCCAC	CCCCCCTGAC	840	
Qy	1140	cact	1143		
Db	841	CACT	844		
RESULT	7	AL562604	LOCUS	AL562604	LOCUS
DEFINITION	AL562604	lTL_NFL003_NBC3	853 bp	mRNA	linear
ACCESSION	AL562604	prime, mRNA	sequence.		
VERSION	AL562604.1	GI:12911188			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Eutheria; Chordata; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	1. (bases 1 to 853)				
JOURNAL	L.I.W.B., Gruber, C., Jesse, J., and Polayes, D.				
COMMENT	Full-length cDNA libraries and normalization				
FEATURES	Unpublished (2001)				
SOURCE	Contact: Genoscope				
	Genoscope - Centre National de Sequencage				
	BP 191 91006 EVRY cedex - France				
	Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
	Location/Qualifiers				
	1..853				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0DC015YJ08"				
	/clone_lib="lTL_NFL003_NBC3"				
	/sex="male"				
	/tissue_type="neuroblastoma cells"				
	/lab_host="DH10B"				
	/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA				
	was primed with a NotI-Oligo(dt) primer. Five prime end				
	enriched, double-stranded cDNA was digested with Not I and				
	cloned into the Not I and Eco RV sites of the PCMVSPORT 6				
	vector. Library was normalized. Library was constructed				
	by Life Technologies. Contact : Feng Liang Life				
	Technologies, a division of Invitrogen 9800 Medical Center				
	Drive Rockville, Maryland 20850, USA Fax : (1) 301 610				
	8371 Email : fliang@lifetech.com URL :				
	http://fulllength.invitrogen.com"				
BASE COUNT	184 a	216 c	238 g	193 t	22 others
ORIGIN					



```

|||||
Db 350 GGTGGCCCGCTGCRGTGAGGGGCCACACCTCGTGGCCGGAATGAGCAACTATCT 291
QY 1010 -gcgaagttgagcgacacacgctggtgtccagacagcgagcctgacctatagca 1068
Db 290 CGGAGAGTGGCCGACACACTGCTGTGTTCCAGCACCCGAGCCGACCTCTATGCA 231
QY 1069 atctgggtacccacgctcttaagacttgactctgtgagttgagacctccctccac 1128
Db 230 ATCTGGGTACCCACGTCTTAAGACTTGACTCTGTGATTTAGACCCGCCSCCSAC 171
QY 1129 cccacacccctacctaagactagttgctcaagtgaacttgaggaagatccagacagc 1188
Db 170 CCCCACCCCGCCATGAGACTAGTCTGCTCAGTGTGAGCTTGAGGAGGAGATCCAGGACAGC 111
QY 1189 tccctcaacttgagaacacttgactctctgtatgtaaacacagatgggctgtgggaag 1248
Db 110 TCCCAAACTTGAGAACCTTGACTCTTGTAAGGAACAGATGGGCCGCTTGGGAAG 51
QY 1249 aaacttaacctgagcttaacctgaggtgacagctgacgt 1288
Db 50 AAACCTTCMCTGAGCTTCACCTGAGGTGAGACTGCACTT 11

RESULT 9
AL563355/c 916 bp mRNA linear EST 16-FEB-2001
LOCUS AL563355 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YM06 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563355
VERSION AL563355.1 GI:12912667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 916)
Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
Location/Qualifiers
1..916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YM06"
/clone_1lb="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact: Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 227 a 230 c 239 g 202 t 18 others
ORIGIN
Query Match 53.6%; Score 759.6; Db 9; Length 916;
Best Local Similarity 96.3%; Freq. No. 2,5e-118;
Matches 775; Conservative 9; Mismatches 20; Indels 1; Gaps 1;
QY 589 agacacatcgtgtgagtagatccagggcgagggcctatgtgaaacaccta 648
|||||

```

```

Db 916 ARACATCTTGTGAGAGTAGAGATTCCAGGGCCCTATGTGTGAAGCAACTCTA 857
QY 649 ccaatgctgggcccaggtcttgatcaacctgacgaacacagcgaagcaaatggtctag 708
Db 856 CCATGCTGGGCCAGCTTAACTTAACTGACATGACACACCCAGGCAAAATTTGTTAG 797
QY 709 ctgtctgctatgacatgacgtctccctgaactctctctgagattgggtcgaagtggagcag 768
Db 796 CTCTCTGCTATMACTATGC -GTTCCCTGMACTCTCTCTGCAATTGCTCAAGCTGGAACAG 738
QY 769 agatacttaactlaacctcagctcttggatccatlaacagccacccacttggaggtgt 828
Db 737 AGATACCTTACTTATTCCTTACGCTTTTGGATCCATTAACAGGCCAGGCCACTGTGGAGTGT 678
QY 829 tgcctgggcccgtgtctatgaaacccagatgctatgtatgtgcagcagcaagtgtgac 888
Db 677 TGCTGCGGGCCCGGCTATGAAACCCAGTGTATGATGAGGCGACAGCAAGTGTGAC 618
QY 889 gccacacatgagaagagagatattatgccaagacatggtgtgagacctctggggagacag 948
Db 617 GCCACCATGGAAGAGAGCAAGTTATGCTTGTGATGTTGATGACCCCTGGGGAACAG 558
QY 949 tggtygccgctgctctgaggggccaagcctctgaccttgccgaatatagacctcaactatc 1008
Db 557 TGGTGGCCCGCTGCTGTGAGAGGGGCCAGCCCTGCTGCGCCGATAGACTCACTTTC 498
QY 1009 tgcgaagttggcgcgacacactgctgtgttccaagacgcaagcctgacctatagca 1068
Db 497 TGGGACAGTTGGCGCGACACCTGCTGTGTCCAGCACCCAGCCCTGACTTATATGCA 438
QY 1069 atctgggtacccacacgtctcttaagacttgactctctgtgagtttgaagacctccctccac 1128
Db 437 ATCTGGGTACCCACCTCTTAAAGACTTACTTGTGAGTTTGAAGCTTGCCCTCCSSC 378
QY 1129 cccacacctgcacatagactagctatgctatgtaacttgtaggaagcaagatccagcacagc 1188
Db 377 CCCBCCCTGCACCTTGTGAGCTAGGGCTCATGTGACTTGGAGGAGAGATCCAGGACACGC 318
QY 1189 tccctcaacttgagaacacttgactctctgtatgtaaacacagatgggctgtgggaag 1248
Db 317 TCCCSNNBTTGGGAACCTTGTACTTCTTGATGAACACAGATGGGCTGTTGGGAAG 258
QY 1249 aaacttaacctgagctgaacctgaggtgagcagcgaggttgagaaggtggaatttat 1308
Db 257 AACTTTTCACCTGAGCTTACCTGAGGTGAGACTGCAAGTTTCAAAAAGTGAATTTTAT 198
QY 1309 atagtcattgttatctcatggaacacgaagttctgctgtaggggctgtagcagcactggcat 1368
Db 197 ATATGCTATGTTTATTTATTCATGGAACACTGAAGTTCTGTGAGGGCTGAGCAGNACTGGCAT 138
QY 1369 tgaataataataataataaagtc 1393
Db 137 TGAATAATATATATATATCATTAAGTC 113

RESULT 10
BI769604 846 bp mRNA linear EST 25-SEP-2001
LOCUS BI769604 603054965F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5',
DEFINITION mRNA sequence.
ACCESSION BI769604
VERSION BI769604.1 GI:15761182
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 846)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLNL1512 row: j column: 22  
High quality sequence stop: 838.

## FEATURES

source

Location/Qualifiers  
1. 846

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_1lib="IMAGE:5204469"

/clone\_1lb="NIR\_MGC\_122"

/lab\_host="DH10B"

/note="Organ: pooled lung and spleen; Vector: pCMV-Sport6;  
Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source

anonymous pool of 24 week female lung, 16 week female

spleen, and 20-22 week male spleens. Library is oligo-dT

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.4 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research genetics tracking code 026. Note:

this is a NIR\_MGC Library."

BASE COUNT 179 a 240 c 226 g 201 t

ORIGIN

Query Match

Best Local Similarity 53.5%; Score 758; DB 10; Length 846;

Matches 843; Conservative 0; Mismatches 0; Indels 35; Gaps 5;

Query 20 tctgggtcagaccgctcccgatcgagccctggaatggttgatcatatctcag 79

Db 2 tctgggtcagaccgctcccgatcgagccctggaatggttgatcatatctcag 60

Query 80 taagaaactactccctacccgctgcgcgcgctgagctcattacacagagctccacag 139

Db 61 -----GCTGGGCTTCATCACCAGGCTCTCTCACAG 90

Query 140 attctgtccctctgtgtctcgtgacctcgagatacctcaactctagttctgtctca 139

Db 91 attctgtccctctgtgtctcgtgacctcgagatacctcaactctagttctgtctca 150

Query 200 gcccaagcccaagagcattgctatctcctctctctcctcgtgaaactccctgtgtgctg 259

Db 151 gcccaagcccaagagcattgctatctcctctctcctcgtgaaactccctgtgtgctg 210

Query 260 gtgcagagtaacatcgacgcagacagacaagaacttaaacatgtgtcagtcagtcgt 319

Db 211 gtgcagagtaacatcgacgcagacagacaagaacttaaacatgtgtcagtcagtcgt 270

Query 320 tggagagctgctcagacagctgtgtcgtcgtcgtcctcgtcgtcgtcgtcgtcgtcgt 379

Db 271 tggagagctgctcagacagctgtgtcgtcgtcgtcctcgtcgtcgtcgtcgtcgtcgt 330

Query 380 tggagagctgctcagacagctgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 439

Db 331 tggagagctgctcagacagctgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 390

Query 440 agaatcacaccagctgtcagaggaatgtgactgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 499

Db 391 agaatcacaccagctgtcagaggaatgtgactgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 450

Query 500 gctgtgccaagactgtgagcagactcagaataatctacaatgtcagctgtcgtcgtcgtcgt 559

Db 451 gctgtgccaagactgtgagcagactcagaataatctacaatgtcagctgtcgtcgtcgtcgt 510

Query 560 caaagagagcagtagtggtcagctacagaggaataacatctgtgtgagtagtagagtagccag 619

Db 511 caaagagagcagtagtggtcagctacagaggaataacatctgtgtgagtagtagagtagccag 570

Query 620 gcaagggcctatgtgtgaaagcaactctacatgctcgtggccagctgtgagctacgt 679

Db 571 gcaagggcctatgtgtgaaagcaactctacatgctcgtggccagctgtgagctacgt 630

Query 680 cagacacacagcagcagagatgtgtcagctgtc-tgtatgacatgtcgtccctgaac 738

Db 631 cagacacacagcagcagagatgtgtcagctgtc-tgtatgacatgtcgtccctgaac 690

Query 739 tctctctgcatgtgtcagctgtgagcagagatctacatctcctcagctgtgag 798

Db 691 tctctctgcatgtgtcagctgtgagcagagatctacatctcctcagctgtgag 750

Query 799 ccattacagccagccacactgtgagagtggtcgtcgtggccgtgtcagcagaccag 858

Db 751 ccattacagccagccacactgtgagagtggtcgtcgtggccgtgtcagcagaccag 808

Query 859 gctatgtatgtgag-cagcacagctgtgagcagcaca 895

Db 809 gctatgtatgtgag-cagcacagctgtgagcagcaca 846

## RESULT 11

AL559163 861 bp mRNA linear EST 16-FEB-2001

LOCUS AL559163 LTI\_NFL008.TC2 Homo sapiens cDNA clone CS0DJ014YA17 5

DEFINITION prime, mRNA sequence.

ACCESSION AL559163

VERSION AL559163.1 GI:12904391

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 861)

AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: [segrete@genoscope.cns.fr](mailto:segrete@genoscope.cns.fr) Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).

Location/Qualifiers

1. 861

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_1lib="CS0DJ014YA17"

/clone\_1lb="LTI\_NFL008.TC2"

/sex="male"

/tissue\_type="T cells from T cell leukemia"

/note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact: Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email: [liang@lifetech.com](mailto:liang@lifetech.com) URL : <http://fulllength.invitrogen.com>

BASE COUNT 176 a 229 c 225 g 201 t 30 others

ORIGIN

Query Match

Best Local Similarity 53.5%; Score 758; DB 9; Length 861;

Matches 790; Conservative 29; Mismatches 9; Indels 4; Gaps 4;

Query 110 gctgggtcattacacagcctcctcagatctcgtcctcgtcgtcgtcgtcgtcgtcgtcgt 169

Db 30 gctgggtcattacacagcctcctcagatctcgtcctcgtcgtcgtcgtcgtcgtcgtcgt 89

Query 170 gataccctacitcagctctgtgtcgtcagccagccagacagcagctatctcctc 229

Db 90 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCACAGGCAATGGCTATCTCTC 149  
 Oy 230 ttccctccctggaacatgccccctggtgctgctgtgtccaggtaaacatcgagccagacaaga 289  
 Db 150 TTCCCTCCGCGCAAGTCCGCCCTGTGTGGCTGTGTGCCAGGTAACTATGACGCCAGATAACGA 209  
 Oy 290 acagaactttaaacaatgtgctgagctggtgttcagagaagctgccaagactggtgtcgtcct 349  
 Db 210 AYAGAACTTTTAAACATGTGCTGACCTGGTTTCGAGAGGCTGCCAAGACTGGGCTCGCT 269  
 Oy 350 ggtcttcctgctcgtgaggtatttgtaacttatgtgcaaggagacctgtgagaagatcaccct 409  
 Db 270 GCTTTCCTGCTGCGCTGAGGCAATTTGACTTTATTTGCAAGGAGACCTGTGAGAGAGCTAACACT 329  
 Oy 410 gtctgaacc-actggtgtggaacactttggagaagatacaccagcttgcaggga-agt 467  
 Db 330 GTCTGAACCTACTGSGGTGGGAAACTTTTGGAAAGATACATCCAGCTTCCAGAGGATATGT 389  
 Oy 468 ggaactggtgctgctcgttgggtgtgtttccatgagcgtgtgccaagacttggagacagctcag 527  
 Db 390 GGACCTCTGCTGCTGTTCTTGGTGGTTCCTATGAGCTGAGCTGAGCTGAGCTGAGCTGAG 449  
 Oy 528 aaatctcaatattgtcgaagctgtcgtctgtgaacagcaaaagggagtagtggccaactaaagg 587  
 Db 450 AAAATCTCAATATGTCACAGTGTCTGCTGAACAGCAAGAGGGCAAGTGTGGCCACTTACAGG 509  
 Oy 588 aagacacatctgtgtgacgtatgagattcccaaggagcaggagccatgtgtgaaagcaacct 647  
 Db 510 AAGACACATCTGTGTGTGACGTAGAGATTTCCAGGGGCGAGGGCCCTATGTGTGAAAGCAACTCT 569  
 Oy 648 accatgacctggagccagcttctgtgagctacactgttcaagcaaacacagcagagcaaatgtgtcta 707  
 Db 570 ACCAATGCTGGGGCCAGTCTTGATGACCTGTCACACACACAGCAGCGCAAGATTGGTCTA 629  
 Oy 708 gctgtctgtatgaatgacgtgtcccttccctgaactctctctggcatgtggtccaagctgtgaga 767  
 Db 630 GCTGTCTGTATGACATATCGGTCCCTCCACTCTCTCRAATTTGGCTCMAGCTGGGCCCA 689  
 Oy 768 gagatctacactacactccttcaagcttctgtgatcca-ttcaaggccagcccaacttgg-agg 825  
 Db 690 GAGAACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 749  
 Oy 826 tgttctgtgagggcccggtgtctatcgaaacccaggtgtctatgtatgtgagcagcagcaagtg 885  
 Db 750 TGTGTCTGTGGGGCCGCTGCTCTCMGMAAACCAAGTCTTMMGTATGATKCMGACGACAGTGTG 809  
 Oy 886 gaagccaccatgaagaagaagcaagltatgtgcccacagcaatgtgtgtagacc 937  
 Db 810 GACGCCACATGATGAGAAAGAGAGCAAGTATGTGGCCACAGCATGTGTGTAGACCC 861

RESULT 12  
 Db1822844  
 LOCUS B1822844 928 bp mRNA linear EST 04-OCT-2001  
 DEFINITION 603040166r1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5180897 5',  
 mRNA sequence.  
 ACCESSION B1822844  
 VERSION B1822844.1 GI:15934394  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 928)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [rsraus@nhi.nih.gov](mailto:rsraus@nhi.nih.gov)  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

FEATURES	Source
High quality sequence stop: 822.	
Location/Qualifiers	
1. .928	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:5180897"	
/clone_lib="NIH_MGC_115"	
/lab_host="DH10B"	
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH MGC library."	
BASE COUNT	198 a 256 c 259 g 214 t 1 others
ORIGIN	
Query Match	52.9%; Score 748.6; DB 10; Length 928;
Best Local Similarity	97.3%; Pred. No. 1.7e-116;
Matches 825; Conservative	0; Mismatches 15; Indels 8; Gaps 6;
QY 110	gctggagcttcacaccaggccctccacagatcttcctcctctgctcgtgctcgtgac
DB 41	gctggagcttcacaccaggccctccacagatcttcctcctctgctcgtgac
QY 170	gatacctcaactctcaagtaacttctgtctcagccaggccagaccatgctatctctc
DB 101	gatacctcaactctcaagtaacttctgtctcagccaggccagaccatgctatctctc
QY 230	ttccctctcggagacgcccctctgctgctgtgtgctcaggtgaatctgaagccagaca
DB 161	ttccctctcggagacgcccctctgctgctgtgtgctcaggtgaatctgaagccagaca
QY 290	acagaaactttaaacaatgctcgtcagctggttcgagaggtctgccaactggtgctgct
DB 221	acagaaactttaaacaatgctcgtcagctggttcgagaggtctgccaactggtgctgct
QY 350	ggccttcctcgtcgtgagcaattgacttcaatgacagggacccctgcagagaaagctaac
DB 281	ggccttcctcgtcgtgagcaattgacttcaatgacagggacccctgcagagaaagctaac
QY 410	gctctaaacacacgggtggaagaacatttgaagaataaacaccagcgttcgacgggaatg
DB 341	gctctaaacacacgggtggaagaacatttgaagaataaacaccagcgttcgacgggaatg
QY 470	actcgtgctcgtcctggtgtgttccatgagcgtgtgccaagacttggaagcaactccaga
DB 401	actcgtgctcgtcctggtgtgttccatgagcgtgtgccaagacttggaagcaactccaga
QY 530	aacttacaatgttcaacgtgtcgtcgaacagcaaaagggcagtagtggtgccaattacaga
DB 461	aacttacaatgttcaacgtgtcgtcgaacagcaaaagggcagtagtggtgccaattacaga
QY 590	gacacatcgtgtgacgttagagatccacagggagggagcctatgtgtgaagaactctac
DB 521	gacacatcgtgtgacgttagagatccacagggagggagcctatgtgtgaagaactctac
QY 650	catgctcgtggcccaagctcttgagtcacactgtcagcacacaccagagcaagatgctagc
DB 580	catgctcgtggcccaagctcttgagtcacactgtcagcacacaccagagcaagatgctagc
QY 710	tgctcgtcgtacgtacgtcgttcctcgtgaactctctgtgcatctggtcgaagctgagcaga

Db 640 TGCTGCTATGACATCGGTTCCCTGAACCTCTCTGCGCATGGCTCAAGCTGGACGAGA 699  
Qy 770 gatactacc-tatccctcagctttgtatccatctac--ggcccaagccactggag 825  
|||||  
Db 700 GATACCTACTTATCCTTCAAGCTTTTGATTCATATACAGGGCCAGCCCATCTGGAGG 759  
Qy 826 tttgtctgagggccgctgctatcgaacccagtgctatg-tagtggcagagacagct 884  
|||||  
Db 760 TGTTCCTGGGGCCCTGCTATCGAAACCCAGTCTATGTTAGTGGCGCAGCAGATGT 819  
Qy 885 gg-acgccacacatgagaagagcagttatgcccacagcatgtgtgtagacccctgg 943  
|||||  
Db 820 GGAAACCCACCATGAGAGAGAGC-AGTTATGGGCAACAGCATGTGTAAACCTGGGG 878  
Qy 944 aacagtgg 951  
| | |  
Db 879 ACAGGGGG 886  
RESULT 13  
Bg436916 793 bp mRNA linear EST 14-MAR-2001  
LOCUS 602488424F1 NIH\_MGC\_18 Homo sapiens CDNA clone IMAGE:4620892 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg436916 GI:13343422  
VERSION Bg436916  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 793)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DPH/Gazdar  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1382 row: 0 column: 05  
High quality sequence stop: 767.  
location/Qualifiers  
1. 793  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4620892"  
/clone\_lib="NIH\_MGC\_18"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; CDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."  
BASE COUNT 181 a 210 c 230 g 172 t  
ORIGIN

Query Match 51.2%; Score 724.4; DB 10; Length 793;  
Best Local Similarity 98.0%; Pred. No. 2.1e-112;  
Matches 776; Conservative 0; Mismatches 11; Indels 5; Gaps 4;  
Qy 242 atggccctggggtgtgtgtgcaagtaacatgcagcagcagacagacagactta 301  
|||||  
Db 2 ACTGCCCTGGGTGTGTGTGCGAGGTATGACGACGCGACGACAGCAACAGAACTTAA 61

Qy 302 aacatgtgctgagctgtgttcgagaagggctgcagactgggtgctcgtgcttccctgc 361  
|||||  
Db 62 AACATGTGCTGAGCTGTTCGAGAGGCTGCGACAGCTGGGTGCTGCTGCTGCTGCTG 121  
Qy 362 tgagcattgactcatttcacgcgggaacctgcagagaagctacacctctcgaacct 421  
|||||  
Db 122 TGAGCGATTGTGACTTCTATTCGACAGGAGCCCTGCAGAGAGCTACACTGCTGTAACACT 181  
Qy 422 ggggtggaactcttgagaagatataccagcttgcagaggaatgtgagctcgtgctgc 481  
|||||  
Db 182 GGCTGGGAACCTTTTGGAAGATACACCACTTGGCCAGGGAATGTGACTCTGGCTGTTC 241  
Qy 482 ctggtgtgttccatgagcgtgtgcgaagacttggagagcagactcagaataatctacaat 541  
|||||  
Db 242 CTGGGTGTTTCATGAGAGGCTGCCAAGACTGGAGCAGACTCAGAAATATCTACAAATTG 301  
Qy 542 tcaagtgtcgtcgtgaagaagaaggagcagtagtgccacttaaggaagaacaactctgt 601  
|||||  
Db 302 TCACGTGCTGCTGTAACAGCAAGAGGGGCACTAGTGGCCACTTACAGGAACACATCTGTG 361  
Qy 602 tgaagtagagattccagggcaggggacctatgtgtgaagcaactctacatgctgtggcc 661  
|||||  
Db 362 TGACGTAGAGATTCCAGGGCAGGGGCTTATGTGTGAAGCAACTTACCATGCTTGCGCC 421  
Qy 662 cagctctgagtcacctgtcagcacaccagcaggaagatltgtctagctgtctatga 721  
|||||  
Db 422 CAGCTTTGAGTACACCTGTGACAGCACACAGAGCAAGATTGTTGTTAGTGTGCTATGATGA 481  
Qy 722 catgagctccctgaaactctctcgtgcaatgtgctcaagcttgagagagataactctta 781  
|||||  
Db 482 CATTGCGTTCCCTGAACCTCTCTGCAATGTGGCTCAAGCTGGAGCAGAAATCTTACCTTA 541  
Qy 782 tccctt-cagcttctgtatccattacagggccagcccaagcaggaggtgtgtcctgaggg 839  
|||||  
Db 542 TCTTTACAGCTTTTGGATTCATTACAGGCCCACTGGAGAGTGTGCTGCGGGCCC 601  
Qy 840 cgtgtctatcgaacccagctgtctatgtgtgagcagcagcagctgtgagcagcagctatga 899  
|||||  
Db 602 CGTGCTATGGAACCAAGCTGCTATGATGTGAGCGAGCAGCAGAGTGTGAGCCGCAATGAG 661  
Qy 900 aagagagcaagtatagtgccacagcagcatgtgtgtagagccc--tggggaacaagtgtgtgcc 957  
|||||  
Db 662 AAGAGAGCAAGTATATGCGCACACATGATGTGAGACCCCTGGGGAACAGTGTGGCCCC 721  
Qy 958 gctgctctgaggggccagcctctgctgtg-cccgaatagacctaacctctgagag 1016  
|||||  
Db 722 GCTGCTCTTGAGGGGCCAGGCTGTGCTTGGCCCGAATGACCTCAAGTATCTGGACAG 781  
Qy 1017 ttgagccgagac 1028  
|||||  
Db 782 TTGCGCGCAAC 793  
RESULT 14  
Bg703176 790 bp mRNA linear EST 07-MAY-2001  
LOCUS 602686082F1 NIH\_MGC\_95 Homo sapiens CDNA clone IMAGE:4818405 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg703176 GI:13975252  
VERSION Bg703176  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 790)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10720 row: d column: 22  
 High quality sequence stop: 766.

## FEATURES

SOURCE

Location/Qualifiers

1..790

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4818405"

/clone\_lib="NIH\_MGC.95"

/tissue\_type="hippocampus"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptpr (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag ) ; Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 167 a 226 c 206 g 189 t  
 ORIGIN

Query Match 50.2%; Score 711; DB 10; Length 790;  
 Best Local Similarity 95.2%; Pred. No. 3.8e-110;  
 Matches 777; Conservative 0; Mismatches 6; Indels 33; Gaps 3;

QY 4 cactgcgcgcgcctnttgctccaagccgcctccgagtcgagcctgcgaatgttt 63  
 DB 6 CACTGCGCGCGCGCTTCGCTCCAGACCGCCCTCCGATGGAGCCCGCAATGGTTT 65  
 QY 64 tggctatatctcatgtaagactactccctaccgctgcgcgcgcgcgcgttca 123  
 DB 66 TGGCTATATCTTCAT-----GCTGGGCTTCATTA 94  
 QY 124 ccaagcctccctacagatctctcctctgtgtcttgactccgatactcaactc 183  
 DB 95 CCAGGCGTCCCTACAGATCTCTGCTTCGTGCTGAGTCCGATACCTCAACTCT 154  
 QY 184 cagtaacttctgctcaagcccaagcgcagcatgtctctctctctcctcgaa 243  
 DB 155 CAGTACTTGTGCTCAGCCAGGCCAGGCCATGGCTATCTCTCTCTCGGAAC 214  
 QY 244 tggcctgtgtgctgtgtgcaggttaacatgcagccagaagaagaacttaaa 303  
 DB 215 TGGCCCTGTGCTGTGTGCCAGTACATCAAGCCAGACAGCAAGAACTTAA 274  
 QY 304 catgtctgaagctgttgcagagagctgcagagctgggtgctgctgcttctgct 363  
 DB 275 CATGTGTCAGCTGCTGTCAGAGGCTGCCAGACTGGGTCCTGCTTCCTGCG 334  
 QY 364 aggcattgactcaatgcagagagcctgcagagagcgttaacactgttgaacact 423  
 DB 335 AGGCACTTGTGACTTGTGACCGGAGCCCTGCAGAGACGCTTACACCGCTGA 394  
 QY 424 gtgggaacttttgaagaatacaccagactgcaggaagtgtgactgtgctct 483  
 DB 335 GTGGAAACTTTTGAAGAATACACCACTTGCCAGGGAATGTGACTGTGCTCT 454  
 QY 484 tgggtgttctcatgagcgttgcagagactggagagcagactcagaatactcaat 543  
 DB 455 TGGGTGTTCATGATGATGCGCAAGACTGGAGAGAGACTGAGAAATCTCAAT 514  
 QY 544 acgtctctctgaagcagaaggcagtagtgcacacttaacaggaaacacatctgt 603  
 DB 515 ACATGCTGCTGAAGCAAGGAGGAGTAGTGCCACTTACAGGAAGACACATCTGT 574

QY 604 acgtagagattccagggcaggggcctatgttgaagcaactctacatgcttggccca 663  
 DB 575 ACGTAGAGATTCCAGGCGAGGGGCCCTATGTGAAAGCAACTTACATGCTGGGCC 634  
 QY 664 gtcttgagtaacctgtcagacaccagcaggaagattgtctagctgtcgtcta-tgac 722  
 DB 635 GTCTTGAGTCACTGTGACACACACCAGCAGCAAGATGTGCTAGCTGTCTATTGAC 694  
 QY 723 atggcgttccctgaactctctctgtgcatgtgtcgaagctggagcagaagtaact 782  
 DB 695 ATGCGGTCCCTGAACCTCTCTGCGATTGGCTCAAGCGTGAGACAGATTAATT 754  
 QY 783 ccttcagcttltgatacca-ttcaagggccagccca 817  
 DB 755 CCTTCAGCTTTTGATCCATTACAGGCCAGGCCCA 790

RESULT 15  
 A1829952/c  
 LOCUS  
 DEFINITION wj47h07.x1 NCI-CGAP\_Lu19 Homo sapiens CDNA clone IMAGE:2406013 3', similar to TR:076091 076091 NTRILASE HOMOLOG 1.; mRNA sequence.  
 ACCESSION A1829952  
 VERSION A1829952.1 GI:5450623  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 787)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-r@mail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/biopr/image/image.html](http://www.bio.llnl.gov/biopr/image/image.html)  
 Insert Length: 1330 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 457.  
 Location/Qualifiers

## FEATURES

SOURCE

1..787

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2406013"

/clone\_lib="NCI-CGAP\_Lu19"

/tissue\_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"

/dev\_stage="adult"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pT733-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 185 a 200 c 220 g 181 t 1 others  
 ORIGIN

Query Match

49.7%; Score 704.4; DB 9; Length 787;

Best Local Similarity 93.4%; Pred. No. 4.8e-109;  
Matches 735; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
QY 615 ccaaggcaggggctatgttgaaagcaactctaccatgcttgccagcttgagtc 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 787 CCCAGGGCAGGGCCTAATGTTGAAAGCAATTCGCCAGGCTTGGCCCGCTGTGCA 728

QY 675 ccttcagcacacagcaggaatgtctagctgtctatgacatcggttcct 734
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 CCGGTCACTACACCAACGCTAAGTTGCTAGCTGTTCGAATGACATCGGTTCCCTC 668

QY 735 gaactctctcgtcattggtcctaagcttgagcagaatlaactlaactccttcagctt 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 GAACCTCTCTCGGCAATGGCTCAAGCTGGACAGAACTACCTCACTCACTTCAAGCTTTG 608

QY 795 ggaaccattacagggccagccactggaggtgttctgcgggcccgtctatgaaacc 854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 GCATCCATACAGGCCGCCACTGGAGGTTGCTCGGGGCCCGTGTATCGAAAC 548

QY 855 cagtcctatgtatgtgcagcagcagtgltgagccaccatgaaagagagcaagttat 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 CAGTGCATATGATGAGCGCAGGCCAGTGTACGCCACATGAGAAAGAGCCAGTTAT 488

QY 915 ggcacacagcatgtgtgtagaccctggggaaacagtggtgcgcgtgctctgaggggcca 974
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 GGCACACAGCATGTGTAGACCCCTGGGGAAACAGTGTGCGCCGCTGTGAGGGGCCA 428

QY 975 ggcctctgcctcccgatagactaactatctgcagcagtgccgacactgct 1034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 GGCCTCTGCTTGCCCGAATGACCTCAACTATCTGCGACAGTTGCGCGACCTGCT 368

QY 1035 gtgtccagcacagcagcctgcactctatgtgcaatctgtgtccaccactgtctaaagac 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 GTGTTCCAGCACCCAGGCCCTGACCTCTATGCAATCTGGGTCAACCCACTGTTTAAAGAC 308

QY 1095 ttgactctgtgagtttagacctgcccctccacccccacccctgccaactatgagctagtg 1154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 TTGACTTCTGTGAGTTTAGACNTGCCCCCTCCACCCCACTGCGCACTATGAGCTAGTG 248

QY 1155 ctcatgtgacttgagagcagatctcagcagacagctccctcaacttggaaccttgaactc 1214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 CTCATGTGACTTGGAGGAGGATCCAGGCACAGCTCCCCCTCACTTGGAGAACCTTGAAGCTC 188

QY 1215 tcttgatggaacacagatgggtgtcttgagaagaanaacttcacctgagcttcaactgag 1274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 TTTTGATGGAGACACAGATGGGCTGCTTGGGAAGAACTTCACTGAGCTTCACCTGAG 128

QY 1275 gtccagactgcagtttcagaaagtgaaatttatatagtcattgttatttcaatgaaac 1334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 GTCAGACTGCAAGTTTCAGAAAGGTGGAATTTATATAGTCATGTGTTTTCATGGAAC 68

QY 1335 tgaagttctgtgagggcttgagcagccactgcatgaaaaataataataataaagtc 1394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 TGAAGTTCTGCTGAGGGCTGAGCAGCCCTGGCATTGAAAAATATATATATCATTAAGTCT 8

QY 1395 aaaaaa 1401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 GAAAAA 1
```

Search completed: April 28, 2002, 04:17:32  
Job time: 10007 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2002, 01:43:40 ; Search time 52.84 Seconds

(without alignments)  
6562.457 Million cell updates/sec

Title: US-09-357-675C-1

Perfect score: 1416

Sequence: 1 gccacacgcgtgcgcctnt.....aaaaaaaaaaaaaaaaaaaaa 1416

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PTCUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	3.3	7218	1	US-08-232-463-14 Sequence 14, Appl
2	41.2	2.9	1332	4	US-09-333-423-1 Sequence 1, Appl
3	41.2	2.9	1602	1	US-08-530-950-3 Sequence 3, Appl
4	41.2	2.9	1602	3	US-08-888-429A-3 Sequence 3, Appl
5	41.2	2.9	1602	4	US-09-149-879-3 Sequence 3, Appl
6	40.4	2.9	1505	4	US-09-385-982-262 Sequence 262, App
7	40.4	2.9	1505	2	US-08-909-965C-13 Sequence 13, Appl
8	39.8	2.8	10660	2	US-08-267-803B-8 Sequence 8, Appl
9	39.8	2.8	10660	4	US-09-041-886-16 Sequence 16, Appl
10	39.2	2.8	1721	5	PCR-US96-00994-3 Sequence 4, Appl
11	39.2	2.8	3982	3	US-08-947-823-4 Sequence 4, Appl
12	38.8	2.7	208	1	US-08-686-878A-37 Sequence 37, Appl
13	38.8	2.7	208	4	US-09-175-928-37 Sequence 37, Appl
14	38.8	2.7	911	2	US-08-924-759-9 Sequence 9, Appl
15	38.8	2.7	911	3	US-09-248-335-9 Sequence 9, Appl
16	38.8	2.7	1393	1	US-07-602-824A-1 Sequence 1, Appl
17	38.8	2.7	1393	1	US-07-983-451-1 Sequence 1, Appl
18	38.8	2.7	1393	1	US-08-261-577-6 Sequence 1, Appl
19	38.6	2.7	2550	6	5258287-23 Patent No. 5258287
20	38.2	2.7	2852	3	US-09-027-137-2 Patent No. 5169941
21	37.8	2.7	1215	6	5169941-7 Patent No. 5169941
22	37.8	2.7	1998	4	US-09-232-200-68 Sequence 68, Appl
23	37.8	2.7	1998	4	US-09-232-197-68 Sequence 68, Appl
24	37.8	2.7	1998	4	US-09-232-201-68 Sequence 68, Appl
25	37.8	2.7	2087	4	US-09-232-191-6 Sequence 6, Appl
26	37.8	2.7	2087	4	US-09-232-200-6 Sequence 6, Appl
27	37.8	2.7	2087	4	US-09-232-197-6 Sequence 6, Appl

28	37.8	2.7	2087	4	US-09-232-201-6 Sequence 6, Appl
29	37.6	2.7	1910	4	US-08-974-691-7 Sequence 7, Appl
30	37.6	2.7	2239	4	US-09-196-390-1 Sequence 1, Appl
31	37.4	2.6	308	2	US-08-721-488-4 Sequence 4, Appl
32	37.4	2.6	536	1	US-08-341-568-1 Sequence 1, Appl
33	37.4	2.6	536	2	US-08-911-020-1 Sequence 1, Appl
34	37.4	2.6	2589	4	US-08-569-749-1 Sequence 1, Appl
35	37.4	2.6	2589	5	PCR-US96-12860-1 Sequence 1, Appl
36	37.2	2.6	90	1	US-08-677-944-1 Sequence 1, Appl
37	37.2	2.6	90	4	US-09-254-048A-1 Sequence 1, Appl
38	37.2	2.6	991	3	US-08-924-747-25 Sequence 25, Appl
39	37.2	2.6	991	4	US-09-247-373B-25 Sequence 25, Appl
40	37.2	2.6	991	4	US-09-296-715-25 Sequence 25, Appl
41	37	2.6	570	1	US-07-885-970A-10 Sequence 10, Appl
42	37	2.6	570	1	US-08-298-687A-10 Sequence 10, Appl
43	37	2.6	570	1	US-08-298-829-10 Sequence 10, Appl
44	37	2.6	599	4	US-09-328-111-147 Sequence 147, App
45	37	2.6	609	1	US-08-530-797-9 Sequence 9, Appl

# ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pt-Fls  
US-08-232-463-14

[illegible]

```

RESULT 2
: Sequence 1, Application US/09333423
: Patent No. 6265636
: GENERAL INFORMATION:
: APPLICANT: Randall, Douglas
: APPLICANT: Thelen, Jay
: APPLICANT: Miernyk, Jan
: APPLICANT: Muszynski, Michael
: APPLICANT: Sewall, Vincent
: TITLE OF INVENTION: Pyruvate Dehydrogenase Kinase
: FILE REFERENCE: Polynucleotides, Polypeptides and Uses Thereof
: CURRENT APPLICATION NUMBER: US/09/333,423
: CURRENT FILING DATE: 1999-06-15
: EARLIER APPLICATION NUMBER: 60/089,998
: EARLIER FILING DATE: 1998-06-19
: NUMBER OF SEQ. ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 1332
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (35)...(1095)
: US-09-333-423-1

```

Query Match	2.98	Pred. 41.2	DB 4	Length 1332
Best Local Similarity	64.98	Pred. No. 0.014		
Matches 61	Conservative 0	Mismatches 33	Indels 0	Gaps 0

RESULT 3  
US-08-530-950-3  
; Sequence 3, Application US/08530950  
; Patent No. 5736381  
; GENERAL INFORMATION:  
; APPLICANT: Davis, Roger J.

```

1  APPLICANT:  Raingeaud, Joel
2  APPLICANT:  Gupta, Shashi
3  APPLICANT:  Derifard, Benoit
4  TITLE OF INVENTION:  CYTOKINE-, STRESS-, AND
5  TITLE OF INVENTION:  ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
6  TITLE OF INVENTION:  KINASES
7  NUMBER OF SEQUENCES:  16
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE:  Fish & Richardson P.C.
10 STREET:  225 Franklin Street
11 CITY:  Boston
12 STATE:  MA
13 COUNTRY:  USA
14 ZIP:  02110-2804
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE:  Floppy disk
17 COMPUTER:  IBM PC compatible
18 OPERATING SYSTEM:  PC-DOS/MS-DOS
19 SOFTWARE:  Patentin Release #1.0, Version #1.30
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER:  US/08/530,950
22 FILING DATE:  19-SEP-1995
23 CLASSIFICATION:  435
24 ATTORNEY/AGENT INFORMATION:
25 NAME:  Fasse, J. Peter
26 REGISTRATION NUMBER:  32,983
27 REFERENCE/DOCKET NUMBER:  07917/010001
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE:  617/542-5070
30 TELEFAX:  617/542-8906
31 TELEX:  200154
32 INFORMATION FOR SEQ ID NO:  3:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH:  1602 base pairs
35 TYPE:  nucleic acid
36 STRANDEDNESS:  single
37 TOPOLOGY:  linear
38
39 US-08-530-950-3

```

Query Match	2.9%	Score 41.2	DB 1	Length 1602
Best Local Similarity	57.9%	Pred. No. 0.016		
Matches 73	Conservative	0	Mismatches 53	Indels 0
			Gaps	0

  

QY	1291	agaagagtggaatttatatagtcaclygttatlucalggaacgtgaagtcgtgag	1350
Db	1444	AGATGAGTGAATTTATGATTAAGGCTTAGGACCTTCAAAAGTGATTAATATTATGATG	1503
QY	1351	gctggagcgacgcctgcattgaaaaatataatcaatcaagaagtcasaaaaaaaaaaaaa	1410
Db	1504	TGTCATATGAGACTCTCCAAAAAaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	1563
QY	1411	aaaaaa	1416
Db	1564	AAAAAA	1569

RESULT 4  
US-08-888-429A-3  
Sequence 3, Application US/08888429A  
Patent No. 6136596  
GENERAL INFORMATION:  
APPLICANT: Davis, Roger J.  
APPLICANT: Whitmarsh, Alan  
APPLICANT: Tournier, Cathy  
TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-  
TITLE OF INVENTION: ACTIVATED HUMAN PROTEIN KINASE KINASES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P. C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA

```
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/888,429A
FILING DATE: 07-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/530,950
FILING DATE: 19-SEP-1995
APPLICATION NUMBER: 08/446,083
FILING DATE: 19-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/053001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 299354
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1602 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 244...1245
US-08-888-429A-3
```

```
Query Match          2.9%; Score 41.2; DB 3; Length 1602;
Best Local Similarity 57.9%; Pred. No. 0.016;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1291 agaaagtggaatttattatagtcattgttattcatggaactgaagtcgctgag 1350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1444 AGATGATGATTAATGATTAAGCTTAGACTTCAAAAGTGATTAATTTAATGATG 1503

QY 1351 gctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1504 TGTCATATGATCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1563

QY 1411 aaaaaa 1416
    |||||
DB 1564 AAAAAA 1569

RESULT 5
US-09-149-879-3
; Sequence 3, Application US/09149879
; Patent No. 6174676
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derjard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/149,879
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/530,950
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1602 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-149-879-3
```

```
Query Match          2.9%; Score 41.2; DB 4; Length 1602;
Best Local Similarity 57.9%; Pred. No. 0.016;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1291 agaaagtggaatttattatagtcattgttattcatggaactgaagtcgctgag 1350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1444 AGATGATGATTAATGATTAAGCTTAGACTTCAAAAGTGATTAATTTAATGATG 1503

QY 1351 gctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1504 TGTCATATGATCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1563

QY 1411 aaaaaa 1416
    |||||
DB 1564 AAAAAA 1569

RESULT 6
US-09-385-982-262/c
; Sequence 262, Application US/09385982
; Patent No. 626234
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 262
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(593)
; OTHER INFORMATION: n = A,T,C or G
```



QY 1286 gtccagaaagtggaatttatatagtcattgtttatcattcagaaactgaagttctgc 1345  
|||||  
Db 10523 GTTTTAAATTAATTTGCTGTATACAGTACAGTTTATTTGTTTCAGTATTACTGTAATA 10582  
QY 1346 tgaaggctgagcagcactgcatgtgaataataataatcataatcaaaataaaaaa 1405  
|||||  
Db 10583 TAAATTAACAGTGCACATTTGCCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 10642  
QY 1406 aaaaaaaaaa 1416  
|||||  
Db 10643 AAAAAAAAAA 10653

RESULT 9  
US-09-041-886-16  
Sequence 16, Application US/09041886  
Patent No. 6235872  
GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
APPLICANT: Rabizadeh, Sharrooz  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of Use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 936..3384  
US-09-041-886-16

Query Match 2.8%; Score 39.8; DB 4; Length 10660;  
Best Local Similarity 56.5%; Pred. No. 0.12; Mismatches 57; Indels 0; Gaps 0;  
Matches 74; Conservative 0;

QY 1286 gtccagaaagtggaatttatatagtcattgtttatcattcagaaactgaagttctgc 1345  
|||||  
Db 10523 GTTTTAAATTAATTTGCTGTATACAGTACAGTTTATTTGTTTCAGTATTACTGTAATA 10582  
QY 1346 tgaaggctgagcagcactgcatgtgaataataataatcataatcaaaataaaaaa 1405  
|||||  
Db 10583 TAAATTAACAGTGCACATTTGCCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 10642  
QY 1406 aaaaaaaaaa 1416

Db 10643 AAAAAAAAAA 10653  
|||||  
RESULT 10  
PCT-US96-00994-3  
Sequence 4, Application US/  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: SELF-RENEWING PLURIPOTENT HEMATOPOIETIC  
TITLE OF INVENTION: STEM CELL COMPOSITIONS, METHODS OF USE,  
TITLE OF INVENTION: AND CULTURE SYSTEMS THEREFOR  
NUMBER OF SEQUENCES: 4  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US SN 08/462,108  
FILING DATE: 05-JUN-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US SN 08/378,144  
FILING DATE: 24-JAN-1995  
Sequence 3, Application US/  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: SELF-RENEWING PLURIPOTENT HEMATOPOIETIC  
TITLE OF INVENTION: STEM CELL COMPOSITIONS, METHODS OF USE,  
TITLE OF INVENTION: AND CULTURE SYSTEMS THEREFOR  
NUMBER OF SEQUENCES: 4  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US UNKNOWN  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US SN 08/462,108  
FILING DATE: 05-JUN-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US SN 08/378,144  
FILING DATE: 24-JAN-1995  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1721 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 63..1436  
PCT-US96-00994-3

Query Match 2.8%; Score 39.2; DB 5; Length 1721;  
Best Local Similarity 66.7%; Pred. No. 0.066;  
Matches 56; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1333 actgaagttctgctgagggctgagcagcactgcatgtgaataataataatcataagt 1392  
|||||  
Db 1637 ATTAAGATCTTTTGTAGGAATTAATAAAGAAAGAAAAAATAAAAAAAAAAAAAA 1696  
QY 1393 caaaaaaaaaaaaaaaaaaaaaa 1416  
|||||  
Db 1697 AAAAAAAAAAAAAAAAAAAAAA 1720

RESULT 11  
US-08-947-823-4

```
; Sequence 4, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Isgouhi
; APPLICANT: Yaghoobi, Jafar
; APPLICANT: Bodeau, John
; APPLICANT: Milligan, Stephen
; TITLE OF INVENTION: Procedures and Materials for Conferring
; TITLE OF INVENTION: Pest Resistance in Plants
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/947,823
; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/18802
; FILING DATE: 09-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,191
; FILING DATE: 10-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-070210US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3982 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 87..3860
; OTHER INFORMATION: /note="Copy 2 CDNA for MI nematode
; US-08-947-823-4
```

```
Query Match 2.8%; Score 39.2; DB 3; Length 3982;
Best Local Similarity 56.1%; Pred. No. 0.1;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
```

```
OY 1280 actgcagttcgaagaagtgaatttatatagtcatttattcatggaactgaag 1339
DB 3851 ATTTAAGTGCATTTGTTGAACCTTGGTGTATATGATTAATAAATATCCG 3910
OY 1340 tctcgtgagggcgtgagcagcctgcatggaataataataataaagcaaaaa 1399
DB 3911 GTGATGAGATTCCTCTTAAGTTCTTTAAACAAAATAATATTTTATATAGTACAAA 3970
OY 1400 aaaaaaaaaaaaaa 1411
DB 3971 AAAAAAAAAAAAAA 3982
```

RESULT 12

```
US-08-686-878A-37
; Sequence 37, Application US/08686878A
; Patent No. 5708157
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallee, Edward
; APPLICANT: Racine, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/686,878A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 208 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-686-878A-37
```

```
Query Match 2.7%; Score 38.8; DB 1; Length 208;
Best Local Similarity 54.0%; Pred. No. 0.028;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
```

```
OY 1304 ttatatagtcattgtttatcatggaactgaagtcctcgtgagggcgtgagcagcact 1363
DB 73 TTATATTATTCCTCCGTAATTTGTCGNGNGATRAATTTGAANATTAACNGCATTAANGNT 132
OY 1364 ggcattgaaataataataataaagtcacaaaaaataaaaaaataaaaaa 1416
DB 133 NATGNTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 185
```

```
RESULT 13
US-09-175-928-37
; Sequence 37, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
```

```
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallee, Edward R.
; APPLICANT: Collins-Racine, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: ML, Sha
```

APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: 6006B.A1172A  
CURRENT APPLICATION NUMBER: US/09/175,928A  
CURRENT FILING DATE: 1998-10-20  
NUMBER OF SEQ ID NOS: 62  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 37  
LENGTH: 208  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (8)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (29)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (31)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (42)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (55)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (65)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (75)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (86)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (91)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (98)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (100)..(101)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (113)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (119)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (121)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (128)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (130)..(131)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (133)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (137)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (161)  
US-09-175-928-37

Query Match 2.7%; Score 38.8; DB 4; Length 208;  
Best Local Similarity 54.0%; Pred. No. 0.028;

Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 1304 ttatatagtcattgttatttcattgaacactgaagttctgcgagcgctgacgacact 1363  
Db 73 ttnattattcccgtaantttgctgngatataatgaanaacngngatgaangnt 132  
QY 1364 ggcattgaaataataataatcacaagtcacaaaaaataaaaaa 1416  
Db 133 natgntaaaaaataaaaaaataaaaaaataaaaaa 185

RESULT 14  
US-08-924-759-9  
Sequence 9, Application US/08924759  
Patent No. 5962229  
GENERAL INFORMATION:  
APPLICANT: MCGONIGLE, BRIAN  
APPLICANT: O'KEEFE, DANIEL  
TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE  
TITLE OF INVENTION: ENZYMES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/924,759  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FLOYD, LINDA AXAMETHY  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: CL-1128  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-892-8112  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 911 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: MAIZE  
IMMEDIATE SOURCE:  
CLONE: CC71SE-B.PK0014.B8  
US-08-924-759-9

Query Match 2.7%; Score 38.8; DB 2; Length 911;  
Best Local Similarity 60.4%; Pred. No. 0.062; Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
QY 1310 taqtcattgttatttcattgaacactgaagttctgcgagcgctgacgacactgacatt 1369  
Db 806 TACTCATGTTTTCATTCCTGCAAGTCACAACACAGTCGCTGCTGTTTACTCTTT 865  
QY 1370 gaaaaataataataataatcacaagtcacaaaaaataaaaaa 1415  
Db 866 TTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 911





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:55:17 ; Search time 21.41 Seconds

(without alignments)  
559.991 Million cell updates/sec

Title: US-09-357-675c-21

Sequence: 1 MLGFTTRPHRFLSLCPGL.....LPVQHRRLDLYGNLGHPLS 327

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	519	29.7	307	1 YJM6_YEAST	P47016 saccharomyc
2	490	28.0	272	1 Y6M1_SYNY3	P55175 synecocyst
3	448.5	25.7	322	1 YAUD_SCHPO	Q10166 schizosacch
4	420	24.0	291	1 Y185_YEAST	P49954 saccharomyc
5	356	20.4	262	1 YBEM_ECOLI	P39874 escherichia
6	353	20.2	262	1 YBEM_ECO57	P58054 escherichia
7	262.5	15.0	340	1 Y480_MYCTU	Q11146 pseudomonas
8	254.5	14.6	285	1 YPOQ_PSEFL	P55176 pseudomonas
9	231	13.2	234	1 YAG5_STRAU	P55178 staphylococ
10	221.5	12.7	346	1 NRJ3_ARATH	P46010 arabidopsi
11	220.5	12.6	261	1 YAG5_STRAU	P55177 staphylococ
12	214.5	12.3	346	1 NRJ1_ARATH	P32961 arabidopsi
13	198.5	11.4	339	1 NRJ2_ARATH	P32962 arabidopsi
14	195	11.2	346	1 ALAM_PSEAE	P11436 pseudomonas
15	179.5	10.3	393	1 BUP_RAT	P03248 rattus norv
16	178.5	10.2	355	1 NRJ4_ARATH	P46011 arabidopsi
17	159	9.1	349	1 NRJ4_TOBAC	Q42965 nicotiana t
18	135	8.9	513	1 YHCK_BACSV	P54608 bacillus t
19	142	8.1	256	1 YAFV_ECOLI	Q47679 escherichia
20	142	8.1	345	1 ALAM_RHOER	Q01360 rhodococcus
21	131	7.5	576	1 NAE2_THEMA	Q9X0Y0 thermotoga
22	115.5	6.6	199	1 Y104_YEAST	P40447 saccharomyc
23	114.5	6.6	679	1 NAE2_MYCTU	P71911 mycobacteri
24	113.5	6.5	356	1 NRJ4_ALCEA	P20960 alcalaigenes
25	112.5	6.4	368	1 CYH7_GLOSO	P32964 gliosocercos
26	108.5	6.2	552	1 NAE2_RHOCA	Q03638 rhodobacter
27	102	5.8	383	1 NRJ1_RHOER	Q02068 rhodococcus
28	102	5.8	567	1 NAE2_AOUAE	Q67091 aquifex aeo
29	100	5.7	365	1 NRJ2_RHOER	Q07021 rhodococcus
30	98.5	5.6	511	1 LNT_PSEAE	Q92186 pseudomonas
31	97.5	5.6	680	1 NAE2_MYCLE	Q9cb66 mycobacteri
32	96	5.5	794	1 SEIL_HUMAN	Q9ubv2 homo sapien
33	94.5	5.4	433	1 FUR4_RAT	Q62994 rattus norv

34	94	5.4	542	1 LNT_CHLTR	O84539 chlamydia t
35	93	5.3	349	1 NRJ5_KLEPO	P10045 klebsiella
36	93	5.3	519	1 LNT_SYNY3	P74055 synecocyst
37	92.5	5.3	520	1 VNN2_HUMAN	Q95498 homo sapien
38	91.5	5.2	541	1 LNT_RHIME	Q52910 rhizobium m
39	91	5.2	512	1 VNN1_MOUSE	Q920K8 mus musculu
40	90	5.1	439	1 LNT_AOUAE	Q67000 aquifex aeo
41	89.5	5.1	1257	1 ERB2_RAT	P06494 rattus norv
42	89	5.1	1255	1 ERB2_HUMAN	P04626 homo sapien
43	88	5.0	790	1 SEIL_MOUSE	Q92296 mus musculu
44	88	5.0	2206	1 POLG_POL3L	P03302 poliovirus
45	87.5	5.0	500	1 VNN3_MOUSE	Q9q225 mus musculu

## ALIGNMENTS

RESULT	1	STANDARD	PRT	307 AA.
YJM6_YEAST				
ID	YJM6_YEAST			
AC	P47016;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	HYPOTHETICAL 34.7 KDA PROTEIN IN SPT10-GCD14 INTERGENIC REGION.			
GN	YJL126W OR J0706.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN=5288C / FY1679;			
RX	MEDLINE=97103775; PubMed=8948101;			
RA	Czapluch C., Kordes E., Pujol A., Jauniaux J.-C.;			
RT	"Sequencing analysis of a 40.2 kb fragment of yeast chromosome X			
RT	reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,			
RT	SPT10, GCD14, RPE1, P086, NCA3, ASF1, CCR7, GZF3, two tRNA genes,			
RT	three remnant delta elements and a Ty4 transposon."			
RL	Yeast 12:1471-1474(1996).			
CC	- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL: Z49401; CAAB9421.1; -.			
DR	SGD: S0003662; YJL126W.			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	InterPro: IPR001110; UPF0012.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPF0012; 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 307 AA; 34693 MW; 48787CA3B10A828E CRC64;			
Query Match 29.7%; Score 519; DB 1; Length 307;				
Best Local Similarity 37.6%; Pred. No. 6e-39;				
Matches 114; Conservative 56; Mismatches 105; Indels 28; Gaps 6;				
QY	45 ELPLVAVCQVTSPPDKOENFTCAELVREARALGACLAFLPEAFDFIARPAETLHL---	101		
DB	4 KIKRVANAQGLSSADLTKNLKLVKELISEALQKADVVFPEASDYSIQPDLMSRYLAQK 63			
QY	102 SEPLGKLEETVYOLARECGIWLISG-GFH--ERGD-WEOTQIKYCHVLNSKGAVVA 157			
DB	64 SPFTIRLOLSSITDVLVDNRSNRINDVSGVHLPESEODLLEGNRVRNVLLYIIDHEGKILQ 123			
QY	158 TYRTHICDVEIFEGQGMCSNSITMPGSLSEPVSTPAGIGIAGVACVDMFPELSLALAQ 217			

```

Db 124 EVOKLELFDVDPNCPILKESKSVOPGKAIPIIEPLKLSALICYDIRPEFSIKLS 183
QY 218 ACAELITVYASAGSTGPAHMEVLLARALETQCYYVAAAGGRH----- 262
Db 184 MCAELITCFPSAFTIKGELARAVDPQCYVLMGQVGMHLSPEMEKOSHMSA 243
QY 263 ---HEKRASGSHMVDPGTVAVARC---SEGPGLCLARIDNLNLRQLRHLPVFQHRP 316
Db 244 LKSSRSREMGSHMVIDPMGKIIAHNDPSTVGPOLILADLRELLQEIKNKPLMNMQRD 303
QY 317 DLY 319
Db 304 DLF 306

RESULT 2
Y601_SYNY3 STANDARD: PRT: 272 AA.
AC P55175;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 30.2 KDA PROTEIN SLI0601.
GN SLI0601.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Kocani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 648 to 928 of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D64002; BA10370.1;
CC InterPro: IPR003010; CN_hydrolase.
CC InterPro: IPR001110; UPF0012.
CC Pfam: PF00795; CN_hydrolase; 1.
CC PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 272 AA; 30191 MW; 7787BA9B2BE03A1 CRC64;

```

```

Query Match 28.0%; Score 490; DB 1; Length 272;
Best Local Similarity 40.2%; Pred. No. 2e-36;
Matches 111; Conservative 46; Mismatches 99; Indels 20; Gaps 7;

```

```

QY 48 LVAQCVTSIPDKQNFCTCAELVREARALGACLAFLPEAFDFIARDPAETLHLEPLG 106
Db 5 LAAALQMTSPRLNTEMLQAEELIDLAVQGAELVGLPENFAFLGN---ETEKLEQATAI 61
QY 107 GALLREYTO-LARECGMLSLSGF-----HERGQMEJQOKIYNCHVILNLSGAVVATPR 160
Db 62 ATATKFLQTMARQVYTLTAGCFEPVAVGAG-----KAYNATYLLIAPNGQELARH 114
QY 161 KTHLCDEVEIPGGPMCESNTPGSLSPV---STPAKIGLAVCYDMFPELSIALAOA 218
Db 115 KHLDFDANVPDNTYWSATVAAQOKY-PPVYHSDSFGMLGSLICDVAFPELXRLSLQ 173
QY 219 GAELITPSAFSGITGPAHMEVLLARALETQCYYVAAAGGRHHEKRASGSHMVDPW 278

```

```

Db 174 GADLVFPAAFAFYATGKDHMQVLLQARAIENTCYIAPAQSGCHYERHRTGHAMIIDPW 233
QY 279 GTVVAARCSGPGCLCLARIDNLNLRQLRHLPVFQHR 314
Db 234 GYLADAGEKPELAIETINPDRKOVROOQPSLQHR 269

RESULT 3
YABU_SCHPO STANDARD: PRT: 322 AA.
ID YABU_SCHPO
AC Q10166;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 35.7 KDA PROTEIN C26A3.11 IN CHROMOSOME I.
GN SPAC26A3.11.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z69240; GAA9334.1;
CC InterPro: IPR003010; CN_hydrolase.
CC InterPro: IPR001110; UPF0012.
CC Pfam: PF00795; CN_hydrolase; 1.
CC PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein.
SQ SEQUENCE 322 AA; 35679 MW; EA6F39B160C7F49F CRC64;

```

```

Query Match 25.7%; Score 448.5; DB 1; Length 322;
Best Local Similarity 32.8%; Pred. No. 1.2e-32;
Matches 108; Conservative 63; Mismatches 123; Indels 35; Gaps 9;

```

```

QY 11 RELSLCPGLR--IFOLSVLCAQPPRAMAIISSSCCLP-----LVAVQVTSIPDKQ 62
Db 4 KFFGLVQKTRSFPSLNCYTR---NIMSVASSL-VKDRARIRIGVLOANKKDSK 59
QY 63 NFKCAELVREARALGACLAFLPEAFDFIARDP-----AETLHSEPLGKLEBY 113
Db 60 NQRLARKLVLEAKNGSNVILPELIFN---SPYGTGYFNOYAEPIEISSP-----SYAL 111
QY 114 TOLARECGMLSLSGFHERGQMEJQOKIYNCHVILNLSGAVVATYRTHLCDVEIPGG 173
Db 112 SSMARDTKTYLFGGSIPEP-----KDGKLYNTAMFDPGKLIAYHRKTHLPDIDIPGV 166
QY 174 PGCESNTPGSLSPVSTPAKIGLAVCYDMFPELSIALQAGAEITVYPSAFSGIT 233
Db 167 SPRESDSLSPGAM-TMWDTEYKFGGLGICYDIRPELAMIAARNGCSMITPGAFNLST 225
QY 234 GPAHMEVLLARALETQCYYVAAAGGRHHEKRASGSHMVDPGTVAVARCSGPGCL 293
Db 226 GLHMEVLLARAIVDENEMVACCAPARMDADYHSMGSHSTVVDPRGKIATYATDERPSIV 285
QY 294 ARIIDLNLRLRHLRPVQHRPDLXGNL 322
Db 286 ADIDPSVMSSTARNVPIYQRRFDVYSEV 314

```

RESULT	4			
ID	YLS8_YEAST	STANDARD;	PRT;	291 AA.
AC	P49954;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	HYPOHETICAL 32.5 KDA PROTEIN YLS8351C.			
DE	YLS8351C OR L9638.5.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
XX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,			
RA	Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,			
RA	Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,			
RA	Johnson D., Johnston L., Langston Y., Latreille P., Le T.,			
RA	Mardis E., Meneses S., Miller N., Nian M., Pauley A., Peluso D.,			
RA	Ridken L., Riles S., Taich A., Trevisan E., Vignati D.,			
RA	Wilcox L., Wolhman P., Vaudin M., Wilson R., Waterson R.,			
RL	Submitted (DSC-1994) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.			
CC	-----			
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	CC the European Bioinformatics Institute. There are no restrictions on its			
CC	CC use by non-profit institutions as long as its content is in no way			
CC	CC modified and this statement is not removed. Usage by and for commercial			
CC	CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	CC or send an email to <a href="mailto:license@lsb-sib.ch">license@lsb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U19102; AAB67751.1; .			
DR	SGD; S0004343; YLS8351C.			
DR	InterPro; IPR003010; CN_hydrolase.			
DR	InterPro; IPR001110; UPF0012.			
DR	Pfam; PF00795; CN_hydrolase; 1.			
DR	PROSITE; PS01227; UPF0012; 1.			
DR	Hypothetical protein.			
SO	SEQUENCE 291 AA; 32549 MW; A813744120088827 CRC64;			
Query Match	24.0%; Score 420; DB 1; Length 291;			
Best Local Similarity	34.5%; Pred. No. 3.6e-30;			
Matches	98; Conservative 53; Mismatches 115; Indels 18; Gaps 7;			
QY	49 VAVCOVT-STPDKQONFTCAELVREAR--LGACIAFLPEAFD-----FIARDPAET 98			
DB	13 VALVQSSSPDKMANLGRAATFIRAKKEQDPTLVLLPECFNSPYSDQE--RKYSYV 70			
QY	99 LHLSPPLGGKLLLEETQLARECGMLSLGGFHERQDMEQDTQIKYCNHVLNSKAAVAT 158			
DB	71 INPKEP--STVOFLSNLANKFILVGGTIPELD--PKTDKIYNTSIFEMDEKLIDK 125			
QY	159 YRKTHLCVETIPGCGPMCEENSTMGPSLESVSPNPACKIGAVYDDMRPELSIALQA 218			
DB	126 HRKHLFLPVDLIPNGISFHESEFLSGEK--STTIDTKYKFGVIGIDYDMKPELAMLARK 184			
QY	219 GAEILTPSAGSITGPAPHWEVLLPARAIETQCVVAAACGGRHHEKRASYSWVDPW 278			
DB	185 GAFAMITYSANNTVYGLPHMHLNARSRAVDNQQVYMLCSPARNLQSSVHAHCHSLVDP 244			
QY	279 GTVAVARCEGPEGLCLARIDLNTLROLRRHLPVFGHRRPDLGNL 322			
DB	245 GKIYVAEGEGEIIYAEILDPEVIESFROAVPLTKORRFDVYSDV 288			

```

ID      YBEM.ECOLI      STANDARD;      PRT;      262 AA.
AD      P39874; P77192;
DT      01-FEB-1995 (Rel. 31, Created)
DT      20-AUG-2001 (Rel. 40, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      HYPOTHETICAL PROTEIN YBEM.
GN      YBEM OR B0625/B0626.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
RN      [12]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12;
RX      MEDLINE=97061202; PubMed=8905232;
RA      Oshima T., Aliba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA      Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA      Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA      Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA      Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA      Yano M., Horiiuchi T.;
RT      "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT      corresponding to the 12.7-28.0 min region on the linkage map.";
RL      DNA Res. 3:137-155(1996).
RN      [13]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12;
RX      Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA      Federpriel N., Hyman R., Kaiman S., Komp C., Kurd O., Lew H., Lin D.,
RA      Nannath A.A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL      submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN      [14]
RP      SEQUENCE OF 1-51 FROM N.A.
RC      STRAIN=K12;
RX      MEDLINE=95075315; PubMed=7984109;
RA      Yamanaoka K., Mitani T., Ogura T., Miki H., Hiraga S.;
RT      "Cloning, sequencing, and characterization of multicopy suppressors
RT      of a muhk mutation in Escherichia coli.";
RL      Mol. Microbiol. 13:301-312(1994).
RN      [15]
RP      SEQUENCE OF 31-78 FROM N.A.
RC      STRAIN=K12 / EMG2;
RA      Robison K., O'Keefe T., Church G.M.;
RL      submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN      [16]
RP      CONCEPTUAL TRANSLATION.
RA      Rudd K.E.;
RL      Unpublished observations (JAN-2000).
CC      -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC      -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE K12 SEQUENCE
CC      DIFFERS FROM THE WILD-TYPE SEQUENCE DUE TO A FRAMESHIFT IN
CC      POSITION 67.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE000167; AAC73726.1; ALT_FRAME.
DR      EMBL; AE000167; AAC73727.1; ALT_FRAME.
DR      EMBL; D90703; BAA35268.1; ALT_FRAME.

```





CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L13334; AAA71975.1; -  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR001110; UPF0012.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS01227; UPF0012; 1.  
 KW Hypothetical protein.  
 FT MONTER 1  
 SO SEQUENCE 234 AA; 26495 MW; C1CBADCA5E1389A7 CRC64;

Query Match 13.2%; Score 231; DB 1; Length 234;  
 Best Local Similarity 31.1%; Pred. No. 1.8e-13;  
 Matches 61; Conservative 27; Mismatches 64; Indels 44; Gaps 6;

QY 142 IYVCHVLNKGAVVAVRTHTLGDVEITPGCGMCESTNMPGPSLESPYSTAG----- 196  
 DB 64 IFNTAFADIDXTGKVINQYDMMLV-----PMLDEPAFLTAGKNVPE 104  
 QY 197 -----KIGLAVCYDMRPPELSLALQAGAILTY-----PSAFGSITGPAMHEVLLRA 244  
 DB 105 TFKLSNKGKVTOMCYDLRPELLRYPARSGATITAFVVAQMPFARLN-----HMQVILKA 159  
 QY 245 RALETQCVVAAAQCGHNEKRASY-GHSMVDPWGTIVARCSEBGLCLARIDLNYLRQ 303  
 DB 160 RALENNMYVIGNCGG--YDGKTQYAGSHVAIINPGEIIELESTREKELVTVIDDAVEQ 217  
 QY 304 LRHRLPYFOHRRDLY 319  
 DB 218 QKRAIPFDSLVPHLY 233

RESULT 10  
 NR\_L3\_ARATH  
 ID NR\_L3\_ARATH STANDARD; PRT; 346 AA.  
 AC P46010; O04909;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, last sequence update)  
 DT 20-AUG-2001 (Rel. 40, last annotation update)  
 DE NITRILASE 3 (EC 3.5.5.1)  
 GN NIT3 OR AT3G44320 OR T10D17.110.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=94294436; PubMed=8022831;  
 RA Bartel B., Fink G.R.;  
 RT "Differential regulation of an auxin-producing nitrilase gene family  
 RT in Arabidopsis thaliana.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=98145459; PubMed=9484465;  
 RA Hillebrand H., Bartling D., Weller E.W.;  
 RT "Structural analysis of the nitr2/nit1/nit3 gene cluster encoding  
 RT nitrilases, enzymes catalyzing the terminal activation step in indole-  
 RT acetic acid biosynthesis in Arabidopsis thaliana.";  
 RL Plant Mol. Biol. 36:89-99(1998).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;

RX MEDLINE=21016720; PubMed=11130713;  
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerg W., Unseid M.,  
 RA Farrantum B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Delsony M., Boutry M., Griwell L.A., Maché R., Puidomenech P.,  
 RA De Simone V., Choise N., Attienave F., Robert C., Brotier P.,  
 RA Wuncker P., Catolico L., Weissenbach J., Saurin W., Queller F.,  
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Wurdemann R., Kranz H., Voss H., Holland N., Brandt P., Nyakatura G.,  
 RA Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,  
 RA Reichelt J., Schafte M., Schoen O., Bariges M., Terol J., Climent J.,  
 RA Navarro P., Collado C., Perez-Perez A., Oltmawelder B., Duchemin D.,  
 RA Cooke R., Laude M., Berger-Lilauro C., Parnelle B., Masuy D.,  
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
 RA Monfort A., Argilou A., Flores M., Ligouri R., Vitale D.,  
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P.,  
 RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Millscher J., Sellers P., Gill J.E., Feldlyum T.V.,  
 RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida T.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:820-822(2000).  
 CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE  
 CC INDOLE-3-ACETIC ACID.  
 CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).  
 CC -1- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U09959; AAA19627.1; -  
 DR EMBL: V07648; CAA68936.2; -  
 DR EMBL: A035865; CAB89000.1; -  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR000132; Nitril\_cyn\_hydrolase.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS00920; NITRIL\_CH1\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CH2\_2; 1.  
 KW Hydrolase; Multigene family.  
 FT ACT\_SITE 186 186 BY SIMILARITY.  
 SO SEQUENCE 346 AA; 38022 MW; 70CFE421547F2B5E CRC64;

Query Match 12.7%; Score 221.5; DB 1; Length 346;  
 Best Local Similarity 27.4%; Pred. No. 2.1e-12;  
 Matches 86; Conservative 44; Mismatches 113; Indels 71; Gaps 14;

QY 56 STPDKQNFKTCALVREARALGACLAFLPEAF-----DFIARDPAET 98  
 DB 41 ATIDRAKEF-----IVEAASAKAKLVLPPEAVIGYPPGFRGLAVGVHNEGREFEN 94  
 QY 99 LHLSE-PLGSKLLEETQLARECGWLWSJGCFHERGQDMQOTKIYCHVYLNKSGAVYA 157  
 DB 95 YHSAIKVPGVEVERLAEAGKNVHLVMGATEKDG-----YTYICTALFFSPQOFLG 148  
 QY 158 IYRKTHLCQVE--IPGCGMCESTNMPGPSLESPYSTAGKIGLAVCYDMRPPELSLAL 215  
 DB 149 KHRKVPSTLERICWGG-----DGSITP-----YVDPIKIGAIQWENRPLLYRTAL 198

```

OY 216 AQAAGAILTYSAGSITGPAAHEVLLRARALETQCYVAAQ-CGR----- 261
DB 199 YAKGEITCAPTADSL-----EMQASMIHINAVGGCFVLSAQFCKRRFPEDYLEND 254
OY 262 -----HHEKRSYGHSMVNDPWGVAVARCS-EGPGICLARIDLNTLRRLPVPFOH-R 314
DB 255 IVDTEKHDPVSGGSGVSIISPLGKVLAGPNYESGLVTADLDGLARAKLVFDVVGHS 314
OY 315 RPDLVGNL---GHP 325
DB 315 KPDIK-NLTVNEHP 327

RESULT 11
YAG5_STAU STANDARD: PRT: 261 AA.
ID YAG5_STAU STANDARD: PRT: 261 AA.
AC P53177;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 29.8 KDA PROTEIN IN AGR OPERON (ONF 5).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE GAL;
RX MEDLINE=96004766; PubMed=7565609;
RA Novick R.P., Projan S.J., Kornblum J., Ross H.F., Ji G.,
RA Kreiswirth B., Vandenesch F., Moghazeh S.;
RT "The agr P2 operon: an autocatalytic sensory transduction system in
RT Staphylococcus aureus."
RL Mol. Gen. Genet. 248:446-458(1995).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52543; CAA36779.1;
DR InterPro: IPR003010; CN_hydrolyase.
DR Pfam: PF00795; CN_hydrolyase; 1.
DR PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein.
SQ SEQUENCE 261 AA: 29826 MW: 2303095A8C18F838 CRC64;

Query Match 12.6%; Score 220.5; DB 1; Length 261;
Best Local Similarity 26.1%; Pred. NO.1.8e-12;
Matches 75; Conservative 53; Mismatches 114; Indels 45; Gaps 12;

OY 45 ELPLV-----AVCVSTPPDKQONFKTCAELVREARIGACIAPLEAFDIAPDET 98
DB 7 QLPYFGDSKNETQITQFEKMN-----AEV-----YVVVLPKMN-----NGYDL 49
OY 99 LHLSPGLGKLEET-----QLAREGIMWLSIGFHERGDMQOTQKIVCHVLSKNAV 155
DB 50 EHLNKNADNINLGOSFSFKHLKEKTKVDIVAGSVN-----IRNOQIFNTAFSVNKSGL 104
OY 156 VATYKRLCLDVEIRPGGWCESNSTM-PGSPLESFVSTPAK-IGLAVCYDMRPPEL 213
DB 105 INEYKVHLV-----PMLAREHEFLTGEVVAEPFQSLSDTYTQGLCYDLRPFELL 156
OY 214 ALAQAAGAILTYSAGSITGPAAHEVLLRARALETQCYVAAQCGRHNEKRAST-GRS 272
DB 157 YPARSGAKIAFYAAQW-PMSRIQLHWHSLKARAENMFEVIGTNSG--FDGNTGYAGHS 213

```

```

OY 273 MVNDPWGVAVARCSGPGICLARIDLNTLRRLPVPFOHRRPDL 319
DB 214 IYINPGLVGLNESADILTVDLNINEVEQORENIPVKSIXLDLY 260

RESULT 12
NR1_ARATH STANDARD: PRT: 346 AA.
ID NR1_ARATH STANDARD: PRT: 346 AA.
AC P32861; Q42543; Q04908;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NITRILASE 1 (EC 3.5.5.1).
GN NIT1 OR AT3G44310 OR T10D17.100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=CV. LANDSBERG ERCTA; TISSUE=Leaf;
RX MEDLINE=92209532; PubMed=1555601;
RA Bartling D., Seedorf M., Mithoefer A., Weller E.W.;
RT "Cloning and expression of an Arabidopsis nitrilase which can convert
RT indole-3-acetonitrile to the plant hormone, indole-3-acetic acid."
RL Eur. J. Biochem. 205:417-424(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Zhou L., Bartel B., Thornburg R.W.;
RT "Nucleotide sequence of the Arabidopsis thaliana nitrilase 1 gene."
RL (In) Plant Gene Register PCR95-130.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98145459; PubMed=9484465;
RA Hillebrand H., Bartling D., Weller E.W.;
RT "Structural analysis of the nitr2/nit1/nit3 gene cluster encoding
RT nitrilases, enzymes catalyzing the terminal activation step in indole-
RT acetic acid biosynthesis in Arabidopsis thaliana."
RL Plant Mol. Biol. 36:89-99(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delany M., Boutry M., Griuell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choisme N., Artiguenave F., Robert C., Brotier P.,
RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Contat A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
RA Reinhardt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Oltmewelder B., Duchemin D.,
RA Cooke R., Lande M., Berger-Liaou C., Burnelle B., Masny D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Monhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo W., Walts A., Uterback T., Fujii C.T., Shea T.P.,
RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Millscher J., Sellers P., Gali J.E., Feldlyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Saito S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsuno M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinozaki S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."

```

RL Nature 408:820-822(2000).  
RN [5]  
RP CHARACTERIZATION.  
RC STRAIN-CV. LANDSBERG ERECTA;  
RX MEDLINE=94286570; PubMed=8016109;  
RA Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;  
RT "Molecular characterization of two cloned nitrilases from *Arabidopsis thaliana*: key enzymes in biosynthesis of the plant hormone indole-3-acetic acid";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).  
CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE INDOLE-3-ACETIC ACID.  
CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, BUT AT A VERY LOW LEVEL DURING THE FRUITING STAGE.  
CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/or-send-an-email-to-license@isb-sib.ch>).  
CC  
DR EMBL; X63445; CAA45041.1; -;  
DR EMBL; U38845; AAB05221.1; -;  
DR EMBL; Y07648; CAA68935.2; -;  
DR EMBL; AL353865; CAB88999.1; -;  
DR PIR; S22398; S22398.  
DR InterPro; IPR003010; CN\_hydrolase.  
DR InterPro; IPR000132; Nitril\_cyn\_hydrolase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase; Multigene family.  
FT ACT\_SITE 186  
FT CONFLICT 312 312 Y->H (IN REF. 2).  
SQ SEQUENCE 346 AA; 38178 MW; 8D4F887CAD1E3C1F CRC64;  
  
Query Match 12.3%; Score 214.5; DB 1; Length 346;  
Best Local Similarity 27.2%; Pred. No. 8.7e-12;  
Matches 84; Conservative 44; Mismatches 104; Indels 77; Gaps 14;  
  
QY 56 STPDQONFKCAELVRAARLACLAIFPEAF-----DFTAR 93  
DB 41 ATIDKAERY-----IVRAASKGALVLPFPEGFIGYRGRFGFLAVGVHNEGDERPK 94  
QY 94 DPATLHLSEPLGKLELEYTOLARECGILSLGFFHERGQDMEQTKIYNCHVLLNS-K 152  
DB 95 YHASAIHVPGEVARLAD---VARKNHVYLWGAIEKEG-----YTLV-CTVLEFSPQ 143  
QY 153 GAVNAVITYKTHLCDE--IPGQPMCESNTPGSPLESVPSTPGKIGLAVCYMRPE 210  
DB 144 GQFLKHKHKLMPSTLERCIWGG---DGSTIP-----VYDTPIGKLGAAICWENRML 193  
QY 211 LSLALAGAGELITVPSAFSGITGPAHMEVLRLARAIEOCVVAAO-CGR----- 261  
DB 194 YRTALYAGIELYCAPRTADGS---KEWSSMLHLAIEGCGVLSAOCFOCRKHPDHD 249  
QY 262 -----HHEKRASYGSHMVPDPCGTVAARCS-EGPGICLARIIDNYLRQLRRHL-P 309  
DB 250 YLFETWYDDKREDSIVSGGSIISPLGQVLAGPNEFESGLTADIDLDIARAKLTGPS 309  
QY 310 VFQARRPDL 318  
DB 310 VGYSRPDV 318

AC P32962; Q96505;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE NITRILASE 2 (EC 3.5.5.1)  
GN NIT2 OR AT3644300 OR T10D17-90.  
OS *Arabidopsis thaliana* (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC Eusterois II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. LANDSBERG ERECTA; TISSUE=Leaf;  
RX MEDLINE=94286570; PubMed=8016109;  
RA Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;  
RT "Molecular characterization of two cloned nitrilases from *Arabidopsis thaliana*: key enzymes in biosynthesis of the plant hormone indole-3-acetic acid";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;  
RX MEDLINE=94294436; PubMed=8022831;  
RA Bartel B., Fink G.R.;  
RT "Differential regulation of an auxin-producing nitrilase gene family in *Arabidopsis thaliana*";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;  
RA Zhou L., Bartel B., Thornburg R.W.;  
RT "Nucleotide sequence of a pathogen induced nitrilase gene from *Arabidopsis thaliana*: Nit2";  
RL (in) Plant Gene Register PGR96-006.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;  
RX MEDLINE=21016720; PubMed=11130713;  
RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unselde M., Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermair B., Dalseny M., Boutry M., Grivell L.A., Mache R., Pligdemoech P., De Simone V., Choisme N., Artiguenave F., Robert C., Bortler P., Wincker P., Catolico L., Weissenbach J., Sautin W., Quetier F., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V., Wummbach E., Dronek H., Erfle H., Jordan N., Bangert S., Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nakamura G., Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B., Conrad A., Hornischer K., Kauer G., Loehmert T.-H., Nordsiek G., Reichelt J., Scharle M., Schoen O., Barques M., Terol J., Clement J., Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D., Cooke R., Laude M., Berger-Liauro C., Purrelle B., Masny D., de Haan M., Maarse A.C., Alcaraz J.-P., Cortet A., Casacuberta E., Monfort A., Argitlon A., Flores M., Liguori R., Vitale D., Mannhaupt G., Haase D., Schott H., Koo H.L., Zaccaria P., Mewes H.-W., Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J., Rooney T.H., Rizzo M., Wais A., Uterback T., Fujii C.Y., Shea T.P., Craey T.H., Haas B., Mafti R., Wu D., Peterson J., Van Aken S., Pal G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V., Pruss D., Lin X., Ntseman W.C., Salzberg S.L., White O., Venter J.C., Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E., Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;  
RT "Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*";  
RL Nature 408:820-822(2000).  
CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE INDOLE-3-ACETIC ACID.  
CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).  
CC -1- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.  
CC -1- DEVELOPMENTAL STAGE: BARELY DETECTABLE IN YOUNG ROSETTES, BUT IS



CC STRONGLY EXPRESSED DURING BOLTING, FLOWERING, AND ESPECIALLY FRUIT  
CC DEVELOPMENT.  
CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X68305; CAA48377.1; -  
DR EMBL: U09958; AAB60275.1; -  
DR EMBL: U38845; AAB05220.1; -  
DR EMBL: AL353865; CAB86998.1; -  
DR PIR: S31969; S31969.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR InterPro: IPR000132; Nitril\_cyn\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
KM Hydrolase; Multigene family.  
FT ACT\_SITE 179 179 BY SIMILARITY.  
FT CONFLICT 37 37 E -> G (IN REF. 3).  
FT CONFLICT 48 48 S -> T (IN REF. 3).  
SQ SEQUENCE 339 AA; 37153 MW; 06CDE76D2FDC24A7 CRC64;

Query Match 11.4%; Score 198.5; DB 1; Length 339;  
Best Local Similarity 25.9%; Pred. No. 2.3e-10;  
Matches 78; Conservative 47; Mismatches 115; Indels 61; Gaps 12;

QY 59 DKQNFKTAEIVREARLACIAPLPEAF-----DPIADPAETLHL 101  
DB 31 DDPATLEKANKRIVEAASGSELVPEAFIGYPRFGILGVGHNEGRDEFRKYHA 90  
QY 102 SE-PIGSKLEIYETOLARCGLSLSGFHERGDMQEQOKIYNCHVILNSKAAVATYR 160  
DB 91 SAIRKPGPEVEKLAELAGNNVYLVMGAIKDG-----YTLCTALFSPQOGPIGKHR 144  
QY 161 KTHLCDVE--IPQGPMSNSTMPPSLESPPSTPAKIGLAVCYDMRPPELSLAQA 218  
DB 145 KLMPTSLERCINGOG-----DGSITP-----YDPTIGKLAICENRNPLYRTALYAK 194  
QY 219 GAELITYPEAFSITGPAHWEVLLARAIETQCYVVAQAQ-CGR-----HH 263  
DB 195 GTIELCAPPADGS---KEMQSMILHAIIEGCGFVLSACQCLRKDFPDHPTVLTWDYD 250  
QY 264 EKR---ASYGSMVVDPMGTYVARCS-EGPGICLARIDLNTYROLRRL-PYFGHRRPD 317  
DB 251 DKEPSIYSGGSVITISPLGQVLAGPNESEGLITADLDGDVARKLYFDSVGHYSRPD 310  
QY 318 L 318  
DB 311 V 311

RESULT 14  
ALAM\_PSEAE STANDARD: PRT; 346 AA.  
AC P11436;  
DT 01-OCT-1989 (Rel. 12. Created)  
DT 20-AUG-2001 (Rel. 40. Last sequence update)  
DT 20-AUG-2001 (Rel. 40. Last annotation update)  
DE ALIPHATIC AMIDASE (EC 3.5.1.4).  
GN AME OR PA3366.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae.  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE.

RC STRAIN-PAC142;  
RX MEDLINE-87219101; PubMed-3108029;  
RA Ambler R.P., Auffret A.D., Clarke P.H.;  
RT "The amino acid sequence of the aliphatic amidase from Pseudomonas  
RT aeruginosa.";  
RL FEBS Lett. 215:285-290(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-87219102; PubMed-3108030;  
RA Brammar W.D., Charles I.G., Matfield M., Liu C.-P., Drew R.E.,  
RA Clarke P.H.;  
RT "The nucleotide sequence of the ame gene of Pseudomonas aeruginosa.";  
RL FEBS Lett. 215:291-294(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01.  
RX MEDLINE-20437337; PubMed-10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizler J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -1- FUNCTION: ALLOWS THE ORGANISM TO GROW ON SHORT CHAIN-LENGTH  
CC ALIPHATIC AMIDES.  
CC -1- FUNCTION: ENABLES THE ORGANISM TO USE ACETAMIDE AS BOTH CARBON  
CC AND A NITROGEN SOURCE.  
CC -1- CATALYTIC ACTIVITY: A MONOCARBOXYLIC ACID AMIDE + H(2)O = A  
CC MONOCARBOXYLATE + NH(3).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M27612; AAA25697.1; -  
DR EMBL: AE004759; AAG06754.1; -  
DR PIR: A26741; A26741.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
KM Hydrolase; Complete proteome.  
FT CONFLICT 40 40 E -> D (IN REF. 1 AND 2).  
FT CONFLICT 169 169 G -> P (IN REF. 2).  
FT CONFLICT 317 317 E -> D (IN REF. 1 AND 2).  
SQ SEQUENCE 346 AA; 38495 MW; E19CEB474EB92B93 CRC64;

Query Match 11.2%; Score 195; DB 1; Length 346;  
Best Local Similarity 25.2%; Pred. No. 4.8e-10;  
Matches 76; Conservative 43; Mismatches 129; Indels 54; Gaps 11;

QY 41 SSSCELPYAVCOVSTPDKQNFKTCAELVREARL-----GACIAPLE-A 87  
DB 7 SSSNDYGVAVVNV-----KMPRLHTAAEVLDAENKIAEMIVGKQGLPGMDLVVPEYS 61  
QY 88 FDIADPAETLHLSEPLGKLEIYEQIARREGIWSLGGFHERGDMQ--TQKIYNC 145  
DB 62 LOGIMYDPAEMMTAAVAIPDETEIFSRACRKANV--VFSLVTGRHEHPKRAYNT 118  
QY 146 HVLNSKGAIVATYRKTHLCDVEIPQGPMSNSTMPPSLESPPSTPAKIGLAVCYD 205  
DB 119 LVLIIDNGEIVQYRKTI-----PWCPIEGWTFGGQTYVSEBPKMKISLITCD 168  
QY 206 MRPELSLAQAQAEILT-----YPSAFGSTITGPAHWEVLLARAIETQCYVVAQAQ 259  
DB 169 GNPPEIWRDCAMKGAEIIVCGGYMPAKDOY-----MMAKAMAMANNNVAAVANA 221

```

OY 260 GRHHKRASVGHSMVDPMGTVVARCSEGP-GLCLARIDLNTLR-----QLRRHLPVFG 312
DB 222 G-FDGVYSFGHSATIGFCGRITLGECEGEEEMGIQVAGLSLQIRARANDQGNHLFKIL 280
OY 313 HR 314
DB 281 HR 282

RESULT 15
BUP_RAT 15
ID BUP_RAT STANDARD: PRT: 393 AA.
AC Q03248;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BETA-UREIDOPROPIONASE (EC 3.5.1.6) (BETA-ALANINE SYNTHASE)
OS (N-CARBAMOYL-BETA-ALANINE AMIDOHYDROLASE).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 202-212.
RC TISSUE=Liver;
RX MEDLINE=9319486; PubMed=8449931;
RA Kvalnes-Krick K.L., Traut T.W.;
RT "Cloning, sequencing, and expression of a cDNA encoding beta-alanine
RT synthase from rat liver.";
RL J. Biol. Chem. 268:5686-5693(1993).
CC -I- CATALYTIC ACTIVITY: N-CARBAMOYL-BETA-ALANINE + H(2)O =
CC BETA-ALANINE + CO(2) + NH(3).
CC -I- COFACTOR: BINDS TWO ZINC IONS.
CC -I- ENZYME REGULATION: ALLOSTERIC ENZYME WITH POSITIVE COOPERATIVITY
CC TOWARD THE SUBSTRATE N-CARBAMOYL-BETA-ALANINE.
CC -I- PATHWAY: THIRD AND FINAL STEP IN THE CATABOLISM OF THE
CC PYRIMIDINE BASES, URACIL OR THYMINE, TO PRODUCE BETA-ALANINE OR
CC 2-METHYL-BETA-ALANINE.
CC -I- SUBUNIT: IN THE ABSENCE OF LIGANDS, THE ENZYME EXISTS AS A STABLE
CC HOMODIMER. ALTHOUGH THIS IS NOT THE MOST ACTIVE FORM OF THE IN
CC ENZYME. THIS NATIVE HEXAMER DISSOCIATES TO AN INACTIVE TRIMER IN
CC RESPONSE TO THE PRODUCT, BETA-ALANINE, OR ASSOCIATES TO THE MORE
CC ACTIVE HOMODIMER IN RESPONSE TO THE SUBSTRATE.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M97662; AAA40804.1; .
DR PIR; S27881; S27881.
DR PIR; A46624; A46624.
DR InterPro; IPR003010; CN_hydrolase.
DR Pfam; PF00795; CN_hydrolase; 1.
KW Hydrolase; Zinc.
KW MOD_RES 71
FT METAL 97 ZINC 1 (POTENTIAL).
FT METAL 101 ZINC 1 (POTENTIAL).
FT METAL 158 ZINC 1 (POTENTIAL).
FT METAL 280 ZINC 2 (POTENTIAL).
FT METAL 293 ZINC 2 (POTENTIAL).
FT METAL 297 ZINC 2 (POTENTIAL).
FT METAL 297 ZINC 2 (POTENTIAL).
SQ SEQUENCE 393 AA; 44042 MW; 723E5AF5B01E3AAD CRC64;

```

```

OY 34 PRAMATSSSCCLPLVAVCOVSTPDKQ---NFKTCAELVREARLGACLAFLPEA--- 87
DB 69 PQLVRGLVQNNIPL-----PTSAPVAEQVSALHKRIEELAEVAAMGVNIICFQEAAM 123
OY 88 -FDGIARDPAETLHUSEPLGKLEETQ-LARECGMLSLGFERGODMDQOTKIYNC 145
DB 124 PFAFCTREKLPWTEFAESAEDGLTTRFCOKLAKKNNM-VVISPILEDRD--HGVLNMT 180
OY 146 HVLLNSKGAVVATYRTHLCDVEIPGQPMCESTNTPGSLSPV-STPAKIGLAVCY 204
DB 181 AVVINSGLVMGKTRKH-----IPRVGDPNISTYMEG-NLGHVFGQTFGRIVANICY 234
OY 205 DMRFPELSIALAQAGAEILTPSAFGSITGPAHWEVLRLARALETQYVVAAGCR--- 261
DB 235 GRHHPLMTLMYSVNGAEIIFNPSATIGELSESMPLEARNMAIANHCFTCALNRVQGEHY 294
OY 262 -----HHEKRASYGHSNVDPMGTVVARCSEG-PGLCLARIDLNTLRQL 304
DB 295 PNEFTSGDGKKAHHDIGYFGSSYVAAPGSRTPGLSRNQDGLVLTMLNLDCQOI 350

```

Search completed: April 26, 2002, 17:02:49  
Job time: 452 sec

Query Match 10.3%; Score 179.5; DB 1; Length 393;  
Best Local Similarity 26.0%; Pred. No. 1.3e-08;  
Matches 77; Conservative 41; Mismatches 139; Indels 39; Gaps 11;



**THIS PAGE BLANK (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:54:42 ; Search time 57.61 Seconds  
(without alignments)  
830.256 Million cell updates/sec

Title: US-09-357-675C-21

Perfect score: 1748

Sequence: 1 MGFTRPPHRLSLCPGL.....LPVFGHRRPDLXGNLGHPLS 327

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1748	100.0	327	4 076091	076091 homo sapien
2	1487	85.1	323	11 088526	088526 mus musculu
3	1484	84.9	323	11 091IN4	091IN4 mus musculu
4	1372	78.5	290	11 09DBF2	09DBF2 mus musculu
5	956	54.7	288	13 091860	091860 xenopus lae
6	710.5	40.6	460	5 076464	076464 drosophila
7	702	40.2	440	5 076463	076463 caenorhabdi
8	645.5	36.9	316	10 091E50	091E50 arabidopsis
9	575.5	32.9	276	3 094660	094660 schizosach
10	537.5	30.7	272	1 027839	027839 methanobact
11	497	28.4	282	2 09HV06	09HV06 pseudomonas
12	485.5	27.8	276	4 09NOR4	09NOR4 homo sapien
13	479	27.4	270	2 083040	083040 plectonema
14	474.5	27.1	276	11 09JH02	09JH02 mus musculu
15	436.5	25.0	275	2 09K0U4	09K0U4 vidrio chol
16	433.5	24.8	283	2 09A9Y3	09A9Y3 caulobacter
17	406	23.2	318	10 09LYH1	09LYH1 arabidopsis
18	401	22.9	270	2 09JTI3	09JTI3 neisseria m
19	398	22.8	270	2 09K0W5	09K0W5 neisseria m

20	395.5	22.6	189	11 09CTG9	09CTG9 mus musculu
21	389.5	22.3	113	6 029367	029367 sus scrofa
22	387	22.1	283	5 09VHE4	09VHE4 drosophila
23	371	21.2	231	11 09D0S1	09D0S1 mus musculu
24	342.5	19.6	264	2 09XA70	09XA70 streptomyces
25	318	18.2	297	2 09RRF7	09RRF7 delinococcus
26	317	18.1	270	1 09HIW8	09HIW8 thermoplasma
27	312	17.8	257	1 030121	030121 archaeoglob
28	291.5	16.7	300	10 09XG19	09XG19 lycopersico
29	279	16.0	287	1 09Y9L1	09Y9L1 aetopyrum p
30	274.5	15.7	271	2 09KE11	09KE11 bacillus ha
31	273.5	15.6	292	2 0916W8	0916W8 pseudomonas
32	264.5	15.1	295	2 09PA02	09PA02 xyella fas
33	262	15.0	270	2 091241	091241 pseudomonas
34	255.5	14.6	259	2 031664	031664 bacillus su
35	253.5	14.5	282	2 09Z505	09Z505 zymomonas m
36	252	14.4	298	12 089413	089413 parametium
37	249.5	14.3	280	2 069808	069808 streptomyces
38	248	14.2	292	2 09AB15	09AB15 caulobacter
39	246.5	14.1	280	2 069939	069939 streptomyces
40	241	13.8	220	10 09Z0H4	09Z0H4 arabidopsis
41	240	13.7	292	2 025452	025452 helicobacte
42	239	13.7	292	2 09Z186	09Z186 helicobacte
43	238.5	13.6	262	2 09L104	09L104 streptomyces
44	233.5	13.4	271	2 09CB47	09CB47 mycobacteri
45	230.5	13.2	294	2 09ZMC7	09ZMC7 helicobacte

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	327 AA.
076091	076091			
AC	076091			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	NITRILASE HOMOLOG 1.			
GN	NIT1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98337986; PubMed=9671749;			
RA	Pekarsky Y., Campigilio M., Siprashvili Z., Druck T., Sedkov Y.,			
RA	Tilib S., Draganescu A., Wermuth P., Rothman J.H., Huebner K.,			
RA	Buchberg A.M., Mazo A., Brenner C., Croce C.M.,			
RT	"Nitrilase and Fhl homologs are encoded as fusion proteins in			
RT	Drosophila melanogaster and Caenorhabditis elegans.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).			
DR	EMBL: AF069984; AAC39901.1;			
DR	EMBL: AF069987; AAC39907.1;			
DR	InterPro: IPR001110; UPF0012.			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPF0012; 1.			
SO	SEQUENCE 327 AA; 35896 MW; 90FTF99D4BA627B1 CRC64;			
Query Match	100.0%; Score 1748; DB 4; Length 327;			
Best Local Similarity	100.0%; Pred. No. 3.2e-150;			
Matches 327; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
OY	1 MGFTRPPHRLSLCPGLIPOLSVLCAPPRMAISSSCCEPIYAVCVSTPDK 60			
DB	1 MGFTRPPHRLSLCPGLIPOLSVLCAPPRMAISSSCCEPIYAVCVSTPDK 60			
OY	61 QONFKTCAELVREARLGCIALFLPAFDFIARDPETHLISPLGKLLSEYTLAREC 120			
DB	61 QONFKTCAELVREARLGCIALFLPAFDFIARDPETHLISPLGKLLSEYTLAREC 120			

```
QY 121 GLMISLGGFHERGQDMEQOTQKLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
    |||
Db 121 GLMISLGGFHERGQDMEQOTQKLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
QY 181 TWPGPSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAAGAILTYPSAAGSITGPAHMEV 240
    |||
Db 181 TWPGPSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAAGAILTYPSAAGSITGPAHMEV 240
QY 241 LIRARAIETOCYVVAACCGRHHERKASYGHSNVDPMGTVVARCSGEGGLCLARIDLNY 300
    |||
Db 241 LIRARAIETOCYVVAACCGRHHERKASYGHSNVDPMGTVVARCSGEGGLCLARIDLNY 300
QY 301 LRQLRRHLPVFQHRRRDLYGNLGHPLS 327
    |||
Db 301 LRQLRRHLPVFQHRRRDLYGNLGHPLS 327
```

```
RESULT 2
088526 PRELIMINARY; PRT; 323 AA.
AC 088526;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE HOMOLOG 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP MEDLINE=98337986; PubMed=9671749;
RX MEDLINE=98337986; PubMed=9671749;
RA Pekarsky Y., Campigillo M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wernuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Fhit homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL; AF069988; AAC40184.1; -.
DR MGD; MGI:1350916; Ntl1.
DR InterPro; IPR003010; CN_hydrolase.
DR InterPro; IPR001110; UPF0012.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA; 35717 MW; 10151CEB151DF2C7 CRC64;
```

```
Query Match 85.1%; Score 1487; DB 11; Length 323;
Best Local Similarity 84.4%; Pred. No. 1.3e-126;
Matches 276; Conservative 27; Mismatches 20; Indels 4; Gaps 2;
```

```
QY 1 MGFTTRPHRLSLLCPELRIPQLSVLCAOPRRPRAAMISSSCSLPLVAVCOVSTPDK 60
    |||
Db 1 MGFTTRPHRLSLLCPELRIPQLSVLCAOPRRPRAAMISSSCSLPLVAVCOVSTPDK 60
QY 61 OONFETCAELVEAARLACLAFLPEAFDFIARDAETLHSEPLGKLEXYTLARCC 120
    |||
Db 61 OONFETCAELVEAARLACLAFLPEAFDFIARDAETLHSEPLGKLEXYTLARCC 120
QY 121 GLMISLGGFHERGQDMEQOTQKLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
    |||
Db 121 GLMISLGGFHERGQDMEQOTQKLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
QY 181 TWPGPSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAAGAILTYPSAAGSITGPAHMEV 240
    |||
Db 181 TWPGPSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAAGAILTYPSAAGSITGPAHMEV 240
QY 241 LIRARAIETOCYVVAACCGRHHERKASYGHSNVDPMGTVVARCSGEGGLCLARIDLNY 300
    |||
Db 241 LIRARAIETOCYVVAACCGRHHERKASYGHSNVDPMGTVVARCSGEGGLCLARIDLNY 300
QY 301 LRQLRRHLPVFQHRRRDLYGNLGHPLS 327
    |||
Db 301 LRQLRRHLPVFQHRRRDLYGNLGHPLS 327
```

```
QY 301 LRQLRRHLPVFQHRRRDLYGNLGHPLS 327
    |||
Db 297 LQOMRQHLPVFQHRRRDLYGSLGHPLS 323
```

```
RESULT 3
Q9R1N4 PRELIMINARY; PRT; 323 AA.
ID Q9R1N4;
AC Q9R1N4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP MEDLINE=98337986; PubMed=9671749;
RX MEDLINE=98337986; PubMed=9671749;
RA Pekarsky Y., Campigillo M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wernuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Fhit homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL; AF069988; AAC40185.1; -.
DR MGD; MGI:1350916; Ntl1.
DR InterPro; IPR003010; CN_hydrolase.
DR InterPro; IPR001110; UPF0012.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA; 35705 MW; FBDD7730713665EF CRC64;
```

```
Query Match 84.9%; Score 1484; DB 11; Length 323;
Best Local Similarity 84.4%; Pred. No. 2.5e-126;
Matches 276; Conservative 26; Mismatches 21; Indels 4; Gaps 2;
```

```
QY 1 MGFTTRPHRLSLLCPELRIPQLSVLCAOPRRPRAAMISSSCSLPLVAVCOVSTPDK 60
    |||
Db 1 MGFTTRPHRLSLLCPELRIPQLSVLCAOPRRPRAAMISSSCSLPLVAVCOVSTPDK 60
QY 61 OONFETCAELVEAARLACLAFLPEAFDFIARDAETLHSEPLGKLEXYTLARCC 120
    |||
Db 61 OONFETCAELVEAARLACLAFLPEAFDFIARDAETLHSEPLGKLEXYTLARCC 120
QY 121 GLMISLGGFHERGQDMEQOTQKLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
    |||
Db 121 GLMISLGGFHERGQDMEQOTQKLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
QY 181 TWPGPSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAAGAILTYPSAAGSITGPAHMEV 240
    |||
Db 181 TWPGPSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAAGAILTYPSAAGSITGPAHMEV 240
QY 241 LIRARAIETOCYVVAACCGRHHERKASYGHSNVDPMGTVVARCSGEGGLCLARIDLNY 300
    |||
Db 241 LIRARAIETOCYVVAACCGRHHERKASYGHSNVDPMGTVVARCSGEGGLCLARIDLNY 300
QY 301 LRQLRRHLPVFQHRRRDLYGNLGHPLS 327
    |||
Db 297 LQOMRQHLPVFQHRRRDLYGSLGHPLS 323
```

```
RESULT 4
Q9DBF2 PRELIMINARY; PRT; 290 AA.
ID Q9DBF2;
AC Q9DBF2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
```

```

GN  NIT1.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C57BL/6J; TISSUE=LIVER;
RX  MEDLINE=21085660; PubMed=11217851;
RA  Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA  Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA  Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA  Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA  Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA  Fledschmann W., Gaasterland T., Glasi C., King B., Kochiwa H.,
RA  Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA  Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA  Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA  Blake J., Boffelli D., Bojunga N., Carninci P., Gariboldi M.F.,
RA  Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA  Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA  Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA  Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA  Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA  Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmung L.,
RA  Wyszynski A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA  Hayashizaki Y.;
RT  Functional annotation of a full-length mouse cDNA collection.;
RL  Nature 409:685-690(2001).
DR  EMBL; AK004988; BAB23723.1; -.
DR  MGD; MGI:1350916; Nt1.
DR  InterPro; IPR003010; CN_hydrolase.
DR  InterPro; IPR001110; UPF0012.
DR  Pfam; PF00795; CN_hydrolase; 1.
DR  PROSITE; PS01227; UPF0012; 1.
SQ  SEQUENCE 290 AA; 31886 MW; 34493DBF2170C71 CRC64;

Query Match      78.5%; Score 1372; DB 11; Length 290;
Best Local Similarity 87.5%; Pred. No. 3e-116;
Matches 252; Conservative 22; Mismatches 14; Indels 0; Gaps 0;

QY  40  SSSCELPLVAVCOVSTPDKOONFKTCAELVREARLGAFLPEAFPIARDPAETL 99
DB  3  SSTSWELPLVAVCOVSTPNKQENFTKCAELVDEARLGAFLPEAFPIARNPETL 62

QY  100 HLESEPLGKLEBYTQLARECGLMISLGFFHERGQDMQETOKIYNCHVLNLSKGAIVAVY 159
DB  63 LLESEPLNGDLGQYSQLARECGIMLSLGFFHERGQDMQENOKIYNCHVLNLSGSAIVASY 122

QY  160 KETHLCDVETPCQPCWESNSTMPGSPLESVPSTPAKIGLAVCYDMRPPELSLATAQAG 219
DB  123 KETHLCDVETPCQPCWESNSTMPGSPLESVPSTPAKIGLAVCYDMRPPELSLATAQAG 182

QY  220 AELTYPSAFGSTGTPAHMEVLLRARAIETOCYVVAACGGRIHERKASGSHMVVDPWG 279
DB  183 AELTYPSAFGSTGTPAHMEVLLRARAIETOCYVVAACGGRIHERKASGSHMVVDPWG 242

QY  280 TVVARSSEPGGLARIDLNTYLRRLRPVQHRPPDLXGNLGHPLS 327
DB  243 TVVARSSEPGGLARIDLNTYLRRLRPVQHRPPDLXGNLGHPLS 290

RESULT 5
Q918G0  PRELIMINARY; PRT: 288 AA.
AC  Q918G0;
DT  01-OCT-2000 (TREMBlrel. 15, Created)
DT  01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT  01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE  NIT1 PROTEIN 1.
GN  NIT1.
OS  Xenopus laevis (African clawed frog).

```

```

OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC  Xenopodinae; Xenopus.
OX  NCBI_TaxID=8355;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RX  PubMed=10959838;
RA  Pace H.C., Hodawadekar S.C., Draganescu A., Huang J., Bieganski P.,
RA  Pekarsky Y., Croce C.M., Brenner C.;
RT  Crystal structure of the worm Nit1pht Rosetta Stone protein reveals a
RT  Nit tetramer binding two Fhit dimers.;
RL  Curr. Biol. 10:907-917(2000).
DR  EMBL; AF284575; AAP87104.1; -.
DR  InterPro; IPR003010; CN_hydrolase.
DR  InterPro; IPR001110; UPF0012.
DR  Pfam; PF00795; CN_hydrolase; 1.
DR  PROSITE; PS01227; UPF0012; 1.
SQ  SEQUENCE 288 AA; 32061 MW; B3673A80609BF042 CRC64;

Query Match      54.7%; Score 956; DB 13; Length 288;
Best Local Similarity 62.7%; Pred. No. 1.3e-78;
Matches 173; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY  47  PLVAVCOVSTPDKOONFKTCAELVREARLGAFLPEAFPIARDPAETLHSEPLG 106
DB  7  PLVAVCOVSTSDKKNFAFCSRLIREAAGRACMFLPEAFPIIGSIEETLSLESLH 66

QY  107 GKLEEYTOIARCGLMISLGFFHERGQDMQETOKIYNCHVLNLSKGAIVAVYRTKTHCD 166
DB  67 GDIQRYTOIARCGLMISLGFFHERGQDMQETOKIYNCHVLNLSKGAIVAVYRTKTHCD 126

QY  167 VEIPGQPCWESNSTMPGSPLESVPSTPAKIGLAVCYDMRPPELSLATAQAGAEILTYP 226
DB  127 VDLQNGVSLRESSTLPGLALIRPITSPPAKIGLGYCDLPPEFSLATAQAGAEILTYP 186

QY  227 SAFGSTGTPAHMEVLLRARAIETOCYVVAACGGRIHERKASGSHMVVDPWGTVARCS 286
DB  187 SAFSTGTPAHMEVLLRARAIETOCYVVAACGGRIHERKASGSHMVVDPWGTVARCS 246

QY  287 EGPGLCLARIDLNTYLRRLRPVQHRPPDLXGNL 322
DB  247 EGPGLCLARIDLNTYLRRLRPVQHRPPDLXGNL 282

RESULT 6
Q76464  PRELIMINARY; PRT: 460 AA.
AC  Q76464;
DT  01-NOV-1998 (TREMBlrel. 08, Created)
DT  01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT  01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE  NITPHIT PROTEIN.
GN  NITPHIT OR CG7067.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

```

RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekullov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacile J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
[2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=98337986; PubMed=9671749;  
RA Pekarsky Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,  
RA Tiliib S., Draganescu A., Wernuth P., Rothman J.H., Huebner K.,  
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;  
RT "Nitrilase and Flit homologs are encoded as fusion proteins in  
RT *Drosophila melanogaster* and *Caenorhabditis elegans*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
DR EMBL: AE003467; AAF47347.1; -;  
DR EMBL: AF069989; AAC39137.1; -;  
DR HSSP: P49789; 2FTT.  
DR FLYBASE: FBgn0024945; NITPHIT.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR InterPro: IPR001310; HIT.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR Pfam: PF01230; HIT; 1.  
DR PROSITE: PS00892; HIT; 1.  
SQ SEQUENCE 460 AA; 52231 MW; 81121A0BC337706 CRC64;

Query Match 40.6%; Score 710.5; DB 5; Length 460;  
Best Local Similarity 48.6%; Pred. No. 3.9e-56;  
Matches 143; Conservative 47; Mismatches 97; Indels 7; Gaps 3;

OY 27 VLCAQPRRAAIISSSELPLVAVCOVTPSPDKQONFKTCALREARLGAACLAFLPE 86  
DB 13 VIAHQQLRRNSVQKQKQOSATIAVGOMKSTSDKANLSQVIELVDRAASQNAQMLFPE 72  
OY 87 AFDFIARPAETLHSEPLGKLELEYQTLARECGMLSLGFGHERGQDWEQTOKTYNCH 146  
DB 73 CDFGSGERTQITELSEGLDGLMAQYRELAKCNKIWISLGVHRRND-----QKIFNAH 127  
OY 147 VLNSKAVAVATYRKTHLCDEVEIPGQGPCESNSTMPGSPLESPEVSTPAKITGLAVCYDM 206  
DB 128 VLNKEGELAAVYRKLIHMFDTTK-EVRLRESDDVTWPGCLERPVSTPGVIGLQICYDL 186  
OY 207 REPPELSTLAAQAGAILLYPSAFSGITGRPAHWEVLLARAATOCYVVAACCGHHNEK 266  
DB 187 RFAEPAVILRLKIGANLITPSAFYATGAHWEILLARAATITQCFVAAAAIGAHNKR 246  
OY 267 ASYGSMAVVDPGTVVARCSEGP-GICLARIIDLNTLRQLRHLPVFOHRPDLX 319  
DB 247 QSMGSHMIVSPMGVNLADCSQEDLDIGTAEVDSLVLQSLYQTMPCFERRRNDIY 300

RESULT 7  
ID 076463 PRELIMINARY; PRT; 440 AA.  
AC 076463  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE NITRILASE AND FRAGILE HISTIDINE TRIAD FUSION PROTEIN NITPHIT.  
GN NITPHIT OR Y56A3A.13  
OS *Caenorhabditis elegans*.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Pelodetelinae; *Caenorhabditis*.  
OX NCBI\_TaxID=6239;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98337986; PubMed=9671749;  
RA Pekarsky Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,  
RA Tiliib S., Draganescu A., Wernuth P., Rothman J.H., Huebner K.,  
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;  
RT "Nitrilase and Flit homologs are encoded as fusion proteins in  
RT *Drosophila melanogaster* and *Caenorhabditis elegans*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
[2]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
none;  
RT "Genome sequence of the nematode *C. elegans*: A platform for  
RT investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL: AF069986; AAC39136.1; -;  
DR EMBL: AL132860; CAB60517.1; -;  
DR HSSP: P49789; 2FTT.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR InterPro: IPR001310; HIT.  
DR InterPro: IPR001110; UPF0012.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR Pfam: PF01230; HIT; 1.  
DR PROSITE: PS00892; HIT; 1.  
DR PROSITE: PS01227; UPF0012; 1.  
SQ SEQUENCE 440 AA; 49936 MW; 34EED23CE92975 CRC64;

Query Match 40.2%; Score 702; DB 5; Length 440;  
Best Local Similarity 50.2%; Pred. No. 2.2e-55;  
Matches 136; Conservative 43; Mismatches 90; Indels 2; Gaps 1;

OY 49 VAVCOVTPSPDKQONFKTCALREARLGAACLAFLPEAFDFIARPAETLHSEPLGK 108  
DB 17 IAVCOVTSNDLEKKNFQAANNIERAGEKKCMVTLPECFDITGLKKNQIDLAATDCE 76  
OY 109 LLEEYQTLARECGMLSLGFGHERGQDWEQTOKTYNCHVNLNSKAVAVATYRKTHLCDE 168  
DB 77 YMEKYRELARKNINWISLGLHKK--DPSDAHPNNTHTLIDSDVTTRAEVYKRLHFLDE 134  
OY 169 IPGQGPCESNSTMPGSPLESPEVSTPAKITGLAVCYDMREPPELSTLAAQAGAILLYPSA 228  
DB 135 IPGKVRHMESEFSKAGTEIIPVDPITGRLGSLICYDVAFPELSLNNKRGQQLSPSA 194  
OY 229 FGSIGPAHWEVLLARAATITQCFVVAACCGHHNEKASYGSMAVVDPGTVVARCSEGP 288  
DB 195 FLNTGLAHWEVLLARAATENOCYVVAACQGAHNPKNQSYGSHAVVDPGVAVACSESR 254  
OY 289 PGLCLARIIDLNTLRQLRHLRVFOHRPDLX 319  
DB 255 VDMCEAIEDLSYVDTLREMQPVFSHRSDLY 285

RESULT 8



09LE50  
ID 09LE50 PRELIMINARY; PRT; 316 AA.  
AC 09LE50  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE NITRILASE 1 LIKE PROTEIN.  
GN AT4G08790.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC Eurosidia II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McCombie R.W., Spiegel L.A., Huang E.N., Nascimento L.U.,  
RA de la Bastide M., Vil D.M., Preston R.R., Matero A., Shah R.,  
RA O'Shaughnessy A., Rodriguez M., Shekher M., Schutz K., See L.H.,  
RA Swaby I., Habermann K., Dedhia N.N., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,  
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,  
RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL161813; CAB82115.1; -;  
DR EMBL: AL161512; CAB78004.1; -;  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
SQ SEQUENCE 316 AA; 34807 MW; 1D740F48DF9A202C CRC64;

Query Match 36.9%; Score 645.5; DB 10; Length 316;  
Best Local Similarity 44.2%; Pred. No. 1.9e-50;  
Matches 129; Conservative 53; Mismatches 91; Indels 19; Gaps 5;

QY 33 RPRAMAISSSCSEPL-----VANCQVSTPDKOONFTCAELVREARLACL 81  
DB 12 KPSLFRITLSSQIPLMATVTKYVAAQMTSVNDLTNFTATCSRLVGEAALAGAKL 71  
QY 82 AFPEAFDFIARPAETLHLEPLGLKLEPYTLARECGMLSLSGFHEGODWEGTOK 141  
DB 72 ICFPENSTFYGDKEGSEVKTAERLDGPMERYCSLARDNSIMLSLGGFERFD-TH 127  
QY 142 IYVCHVLLNSKGAVALVYRTHLCDEVEIPQGPWCESNSTMPGSLSPVSTPAKIGLA 201  
DB 128 LCMTHVVIDDAGIMRIDTYQKMLFEDVDVPGSSSYKSESTFVPGKIVS-VDSPVGRGLT 186  
QY 202 VCDMRFPELSTALA-QAGAEILTPSAFSGSTGPAHEVLLRARALETQCYVAAQCG 260  
DB 187 VCYDLRRPKTYQQLRFEGQAKVLLVPSAFKTYGGEAHEWILLRARALETQCYVAAQAG 246  
QY 261 RHHEKRASTGVVWDPMGTVAVRCSE--GPGCLARIDINYLQLRRLHPV 310  
DB 247 KHHEKRESYDITLIDPMGTIVGRLPDVRVSTGIYVADIDSLDSVTKMPI 298

RESULT 9  
ID 094660 PRELIMINARY; PRT; 276 AA.  
AC 094660  
DT 01-MAY-1999 (TREMblrel. 10, Created)  
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE HYPOTHETICAL 30.4 KDA NITRILASE-LIKE PROTEIN C651.02 IN CHROMOSOME 11.  
GN SPBC651.02.

OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Wood V., Rajandream M.A., Barrell B.G., Volkert G.,  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO THE N-TERMINAL OF NITRILASES.  
DR EMBL: AL035570; CAB37598.1; -;  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
KW Hypothetical protein; Lyase.  
FT ACT\_SITE 156  
SQ SEQUENCE 276 AA; 30421 MW; 07AA741A54297E07 CRC64;

Query Match 32.9%; Score 575.5; DB 3; Length 276;  
Best Local Similarity 44.2%; Pred. No. 3.4e-44;  
Matches 126; Conservative 50; Mismatches 86; Indels 23; Gaps 8;

QY 46 LPLVAVCVSTPDKOONFTCAELVREARLGAFLPAPFPIARDPAETLHL-SEP 104  
DB 1 MTLAAVQOLNSSGSIILKALCKELISQAAKAKCFFPASPFIANSDEAILETNHP 60  
QY 105 LGGKLEBYTLARECGMLSLSGFHEGODWEGTOKIYNCHVLLNSK-----GAVVA 157  
DB 61 DCSKFIDVRESAKRHSIFVNI-CVHE-----PSVKKN-KLLNSLFTLEPLHGETIS 110  
QY 158 TYRTHLCDEVEIPQGPWCESNSTMPGSLSPVSTPAKIGLAVCYDMRPELSTALAQ 217  
DB 111 RYSAHLFDEYKIKNGPPLKESNTTLRGEALIPCKTPGKVGSAICDIPRPEDAIKIRN 170  
QY 218 AGAEILTPSAFSGSTGPAHEVLLRARALETQCYVAAQCGRHHEKRASTGVVWD 277  
DB 171 MGAHIIITPSAFTEKTAHHEVLLRARAALDSQCYVLAPOAGKHNKRASIGSMIVDP 230  
QY 278 WGTVAVARCSE--GP-GLCALARIDINYLQLRRLHPVQHRPDDY 319  
DB 231 WGTVAQYSDISSPNGLIFFADLDLNDVHRTYIFLL--RRNDLY 273

RESULT 10  
ID 027839 PRELIMINARY; PRT; 272 AA.  
AC 027839  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE N-CARBAAMOYL-D-AMINO ACID AMIDOHYDROLASE.  
GN MTH1811.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OC Methanobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DELTA H;  
RX MEDLINE=86037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT delta: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL: AE000934; AAB86277.1; -;  
DR InterPro: IPR003010; CN\_hydrolase.





Query Match	25.0%	Score 436.5;	DB 2,	Length 275;
Best Local Similarity	38.1%	Pred. No. 1.3e-31;		
Matches 104, Conservative	41;	Mismatches 111;	Indels 17;	Gaps 5

Search completed: April 26, 2002, 17:02:22  
Job time: 460 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:52:52 ; Search time 33.36 Seconds  
(without alignments) 746.675 Million cell updates/sec

Title: US-09-357-675C-21

Perfect score: 1748

Sequence: 1 MGFTTRPPHRLSLCPGL.....LPVGHRRPDYGNLHPLS 327

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	702	40.2	440	2	T43198
2	645.5	36.9	316	2	D85088
3	575.5	32.9	276	2	T40601
4	537.5	30.7	272	2	B69109
5	519	29.7	307	2	S56907
6	497	28.4	282	2	B83086
7	490	28.0	272	2	S76524
8	448.5	25.7	322	2	T38399
9	436.5	25.0	275	2	F82325
10	420	24.0	291	2	S51459
11	406	23.2	318	2	T48563
12	401	22.9	270	2	B81834
13	398	22.8	270	2	B81199
14	342.5	19.6	264	2	T36488
15	318	18.2	297	2	F75263
16	312	17.8	257	2	C69264
17	310	17.7	187	2	H64796
18	298	17.0	187	2	D85562
19	279	16.0	287	2	A72454
20	274.5	15.7	271	2	G83780
21	273.5	15.6	292	2	G83608
22	264.5	15.1	295	2	H82556
23	262.5	15.0	340	2	C70743
24	262	15.0	270	2	B83387
25	255.5	14.6	259	2	E69863
26	254.5	14.6	285	2	S58240
27	252	14.4	298	2	T17568
28	249.5	14.3	280	2	T28684
29	246.5	14.1	280	2	T34905

30	241	13.8	220	2	A84673	probable nitrilase
31	240	13.7	292	2	E64614	beta-alanine synth
32	239	13.7	292	2	F71901	hypothetical prote
33	230.5	13.2	294	2	G71949	hypothetical prote
34	229.5	13.1	292	2	E64558	conserved hypothet
35	221.5	12.7	346	2	T49148	nitrilase (EC 3.5.
36	220.5	12.6	261	2	S20793	hypothetical prote
37	218.5	12.5	272	2	T41662	probable nitrilase
38	218	12.5	262	2	C71109	hypothetical prote
39	216.5	12.4	246	2	A70310	conserved hypothet
40	216	12.4	271	2	H83195	conserved hypothet
41	215.5	12.3	346	2	T49147	nitrilase (EC 3.5.
42	214.5	12.3	346	2	S22398	nitrilase (EC 3.5.
43	214	12.2	290	2	B81369	probable hydrolase
44	208.5	11.9	270	2	B84250	hypothetical prote
45	206	11.8	262	2	C75051	hydrolase related

## ALIGNMENTS

RESULT 1  
T43198  
nitrilase/Fhlt protein - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
C/Accession: T43198  
R:Peckarsky, Y.; Campiglo, M.; Siprashvili, Z.; Druck, T.; Sedkov, Y.; Tiliib, S.; Dr  
Proc. Natl. Acad. Sci. U.S.A. 95, 8744-8749, 1998  
A/Title: Nitrilase and Fhlt homologs are encoded as fusion proteins in Drosophila mel  
A/Reference number: 222332; MUID:98337986  
A/Accession: T43198  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-440 <PEK>  
A/Cross-references: EMBL:AF069986; NID:g3228663; PIDN:AAC39136.1; PID:g3228664  
C/Genetics:  
A:Gene: NitPhit

Query Match 40.2%; Score 702; DB 2; Length 440;  
Best Local Similarity 50.2%; Pred. No. 1.9e-54;  
Matches 136; Conservative 43; Mismatches 90; Indels 2; Gaps 1;

QY	49	VAVCVSTPPKQNFCTCAELVREARLGLAFLPEAFDTPARDPAETLHSEPLGK	108
DB	17	IAVQWTSNDLEKFNQAAKNMIFERAGEKCEWFLPECFDFGLKNQIDLAMATDCE	76
QY	109	LEETVQLARECGIMSLGFEHREGDMEQTKIVNCHVLNKGAVVATYRTHCDVE	168
DB	77	YMEKYRELARKNIWLSLGLNHR--DPSDAHPWVTHLIDSGVTRVYKRLHLEDE	134
QY	169	IGGCGPMCESNSTMPGSPLESVPSTPAGKIGLAVCYDMFPELSLAAGAIEILTPSA	228
DB	135	ITGVNLMSESEFSKAGTEHMPDPITIGRLGLSTICDVAFPELSLMNRRKGAQLSFP	194
QY	229	FESITGPAAHEVYLARRALETQCYVAAACGHHHEKRASTYGSWVVDPMGTVAACSE	288
DB	195	FTLNTGLAHMETLRLARRAIENOCYVAAQTGAHNKRQSYGSHSVVDPWGAIVAC	254
QY	289	PGCLARIDLNLRLRLPFGVGHRRPDLX	319
DB	255	VDMCFEIDLSTVDTLREMPVFSHRSDLY	285

RESULT 2  
nitrilase 1 like protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C/Accession: D85088  
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp  
Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.

A:Reference number: A85001; MUID:20083488

A:Accession: D85088

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <STO>

A:Cross-references: GB:NC\_001268; NID:g7267521; PIDN:CAB78004.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4G08790

A:Map position: 4

Query Match 36.9%; Score 645.5; DB 2; Length 316;

Best Local Similarity 44.2%; Pred. No. 1.3e-49;

Matches 129; Conservative 53; Mismatches 91; Indels 19; Gaps 5;

33 RPRMAISSSCELP-----VAVCOVSTPPKQONFKTCAELVREARLGLA 81

12 KPSFTTRITLSSQIPLTWATVTKTVRAAQAQMTSVNDLMTFATCSKLVDEALAGAKL 71

82 AFLPEAFDFIARDPAETLHLESPGLKLEETOLARECGILWLSGFERHGGDMEQTOK 141

72 ICFPENSFVADKDESGSKIAEPLDGPWERYCSLARSNITWLSGFERD-----TH 127

142 IYNCHVLSKGAVALATYRKTLCDELVEIPGQPMCESNSTPPGSLSPVSTPAKIGLA 201

128 LCNHNVVVDAGMIRDTYQKMHLEFDVDPGSSYKESFTVPGRFIVS-VDSVGRGLLT 186

202 VCYDMRPPELSALA-QAGAEILTYPSAFSGITGPAHNEVLIRARAITQCYVAAACG 260

187 VCYDLRFKRIYQOLREFEQAQVLLVPSAFVTGTGEAHWEILLRARAETOCYVIAAAG 246

261 RHHEKRSYGHSMVVDPMGTAVARCS--GPGICLARIIDLVTQLRHHLPV 310

247 KHNEKREYGTGLIIDPMGTIVVGRLPDRVSTGIYVADIDFSLDSVTKMPI 298

RESULT 3

40601

putative nitrilase homolog - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C:Accession: T40601

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Voiclaert, G.

submitted to the EMBL Data Library, February 1999

A:Reference number: Z21940

A:Accession: T40601

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-276 <MCO>

A:Cross-references: EMBL:AL035570; PIDN:CAB37598.1; GSPDB:GN00067; SPDB:SPBC651.02

A:Experimental source: strain 972h; cosmid c651

C:Genetics:

A:Gene: SPDB:SPBC651.02

A:Map position: 2

A:Introns: 49/3; 81/3; 226/3; 232/1

C:Superfamily: hypothetical protein YLR351c

Query Match 32.9%; Score 575.5; DB 2; Length 276;

Best Local Similarity 44.2%; Pred. No. 1.7e-43;

Matches 126; Conservative 50; Mismatches 86; Indels 23; Gaps 8;

45 LPVAVCOVSTPPKQONFKTCAELVREARLGLAFLPEAFDFIARDPAETLHL-SEP 104

1 MTLAVALQNLSSGSLKLAICKELISQAARKGACIFPEASDFTAINSDAEILETHNP 60

105 LGKLLLEETOLARECGILWLSGFERHGGDMEQTOKIYNCHVLSKGAVAL- 157

61 DCSKFIIRVRSATKHSIFVNI-CVHE-----PSKRN--KLNLSSLFIEPLHGLIIS 110

158 TYRTHLCLDVEIPGQPMCESNSTPPGSLSPVSTPAKIGLAVCYDMRPPELSIALAQ 217

Db 111 RYKRAHLFDEIVEIKNGPTLKESNTTLRGEALLPCKTPLGKVSACIPDIRPEQAIKLRN 170

Qy 218 AGAEILTYPSAFSGITGPAHNEVLIRARAITQCYVAAACGRRHHEKRSYGHSMVVD 277

Db 171 MGAHITTPSAFTKTXGAHWEILLRARAIDSQCTVIAPAGCGKNEKRSYGHSMVVD 230

Qy 278 WGTAVARCS--GP-GICLARIIDLNTYLRRLHLPVFOHRRPDL 319

Db 231 WGTAVAGYSDISSPNGLIFADLDLNLVDHVRTYIPDL--RRNDLY 273

RESULT 4

B69109

N-carbamoyl-D-amino acid amidohydrolase - *Methanobacterium thermoautotrophicum* (Stral

C:Species: *Methanobacterium thermoautotrophicum*

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999

C:Accession: B69109

R:Smith, D.R.; Doucellette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.

; Qiu, D.; Spadafora, R.; Viscare, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani,

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: fu

A:Reference number: A69000; MUID:98037514

A:Accession: B69109

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-272 <MTH>

A:Cross-references: GB:AE000934; GB:AE000666; NID:92622924; PIDN:AAB86277.1; PID:9262

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1811

C:Superfamily: hypothetical protein YLR351c

Query Match 30.7%; Score 537.5; DB 2; Length 272;

Best Local Similarity 38.4%; Pred. No. 4e-40;

Matches 109; Conservative 49; Mismatches 95; Indels 31; Gaps 4;

49 VAVCOVSTPPKQONFKTCAELVREARLGLAFLPEAFDFIARDPAETLHLESPGLGK 108

3 LGICOMQYTERKREKNGVAGVIEAERGARLIVLPMF-----TCPIYVN 49

109 LLEETQ-----LARECGILWLSGFERHGGDMEQTOKIYNCHVLSKGAVAL 156

50 LFEYAEEDENCYSTRMSIARELGILHVAQSIP-----EPREGIYNTSVMDGMI 104

Qy 157 ATYKTHCLDVEIPGQPMCESNSTPPGSLSPVSTPAKIGLAVCYDMRPPELSIALA 216

Db 105 GKRRVHLFDINPEISFRSDSLIAGDSV-TVLETQCVWGVGICVDMRPPELSIRMA 163

Qy 217 OAGAEILTYPSAFSGITGPAHNEVLIRARAITQCYVAAACGRRHHEKRSYGHSMVVD 276

Db 164 LGCAEVLLFPGAFNMTTTPAHMRLLVRSKALDNCYCAASPARPSAIVAYGHSMLAD 223

Qy 277 PWGTAVARCSSEPGICLARIIDLNTYLRRLHLPVFOHRRPDL 320

Db 224 PWGSVVCADASPSVITADIDLEAVERIRRLPLLRNRPDYG 267

RESULT 5

S56907

hypothetical protein YJ126w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein J0706

C:Species: *Saccharomyces cerevisiae*

C>Date: 05-May-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Sep-1999

C:Accession: S56907

R:Czaplich, C.; Koides, E.; Pujol, A.; Jauniaux, J.C.

submitted to the Protein Sequence Database, September 1995

A:Reference number: S56891

A:Accession: S56907

A:Molecule type: DNA

A:Residues: 1-307 <CZ1>

A:Cross-references: EMBL:Z49401; NID:g1008323; PIDN:CAA89421.1; PID:g1008324; MIPS:YJ







A: Experimental source: cultivar Columbia; BAC clone F14F18  
C: Genetics:  
A: Map position: 5  
A: Introns: 25/3; 64/3; 139/3; 182/1; 207/1; 233/2; 238/3; 265/3; 293/2  
A: Note: F14F18.210

Query Match	23.28;	Score 406;	DB 2;	Length 318;
Best Local Similarity	33.78;	Pred. No. 2.2e-28;		
Matches 102;	Conservative 45;	Mismatches 116;	Indels 40;	Gaps 7;

```

Oy 46 LPL-----VAVCOVTPSPDQOONFKCAELVYREAALGCLAFPAEPFADRA 96
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 16 LPLAPRLPTKRNICQLQSLVSTSDKKRNISNAKKAIIEPAASKGAKVLLPPTMNSPYNDS 75

Oy 97 ETLHLEPGLGKLLVEYTQIAREBCGLWLSL---GGFHERGQDWEOTOKLYNCHVLLNSK 152
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 76 FPYVAEIEDAGDGASPSTAMLSIEVSKRLKLTIIIGGSIPEBYGD---RLYNCCVFGSD 130

Oy 153 GAVATYRKTKHLCVEIIPGCGPMCESNSTMPGSESP-----VSTPAKGIG 199
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 131 GELAKAKHKHILHPDIDIPKRTTFMESTATLTAAG---EFTLYVDGYNGNLGRLNIIPDGRIG 187

Oy 200 LAVCTDMRFPPLSLAQAQAEIITYSAFGISTGPAHMEVLLRARAIEFQCYVAAAO 259
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 188 IGCIYDINFOELAMIVYARGNHLCYGAENMTGTGPHWELTORARATDQIYV---ATC 244

Oy 260 GRHHEKRASY---GHSWVDPMGTVVVYARCSBGPLCLARIDLVNLRJRLHVFQHRP 316
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 245 SPADSOGAGYAMKHSIYVGFGEVLAATTEHEAIIIAELDIYSLEBRKISLPLNQRRG 304

Oy 317 DLY 319
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 305 DLY 307

```

RESULT 12  
E81834  
conserved hypothetical protein MMA2044 [imported] - Neisseria meningitidis (strain Z2491)  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: E81834  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morello,  
R.; Holtrop, S.; Javelle, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556  
A:Accession: E81834  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-270 <PAR>  
A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85262.1; PID:g738065  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: opaA; MMA2044  
A:Superfamily: hypothetical protein YUR351c

Query Match	22.9%	Score 401:	DB 2:	Length 270:
Best Local Similarity	37.3%	Pred. No. 4.9e-28:		
Matches 103:	Conservative 35:	Mismatches 112:	Indels 26:	Gaps 8:

  

QY	49	VAVCGVSTSPKQONFKCAELVREARLGLACFLPEAPFPIARDPAETLHLSEPTL	CG	107
		: : : : :    : : : : :    : : : : :    : : : : :		
Db	6	VAAQMGVSGVPEPTVAAAMKRLVRAAQGVDWVLLPEYWLGMGNDIDKTLALPEPLGG		65
		: : : : :    : : : : :    : : : : :    : : : : :		
QY	108	KLLEETYLARECGIMLSLGG----	FEHQGDMEQTKIYNCHVLYLNSKGAVAATYRKT	162
		: : : : :    : : : : :    : : : : :    : : : : :		
Db	66	RFQTLSESTAKECGVYL-FGGTVPLQSPDEAG-----	KVNNTLLVYGRDGRKRTGLYHKM	117
		: : : : :    : : : : :    : : : : :    : : : : :		
QY	163	HLCDVEIRGGQPMCSNSTMTPGFS----	LESPYSTRPACRIGLAVCYDMKRFPLSLALQA	218
		: : : : :    : : : : :    : : : : :    : : : : :		
Db	118	HLFGFSGGLGE-RYAAADITLAGEVPHLLAGGVSAAG----	LCYDVRPEP--FEFRQL	169
		: : : : :    : : : : :    : : : : :    : : : : :		

QY	219	GAEIIITYSAGSGSTGPRHVVII	RRARIEPQSVVAAAOGRNKEK	RASVGHSMVDPM	278
		:::	:	:	
Db	170	PRDVLMTPRAFTHTTKGKHWEL	LRAKAVENQSVVAAAOGLHENG	RRTFGHMTIDPM	229
QY	279	STVAVARCSGEPGLCLARDLML	YQOLRNHPVFOHR	314	
				:::	
Db	230	GDVLDLPEEGEGIVADIDARL	LSVNRKRIIPALKYR	265	

RESULT 13  
B81199  
nitrilase NMB041 [imported] - *Neisseria meningitidis* (strain MC58 serogroup B)  
C:Species: *Neisseria meningitidis*  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: B81199  
R:Reteljin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; Li, H.; Qin, H.; Yamathavan, J.; Gill, J.; Scariato, V.; Massignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunli, R.; A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
A:Reference numbers: A81000; MUID:20175755  
A:Accession: B81199  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-270 <TEF>  
A:Cross-references: GB:AE002400; GB:AE002098; NID:g7225659; PIDN:AAF40879.1; PID:g7222222  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB041  
C:Superfamily: hypothetical protein YLR351c

Query Match	22.8%	Score 398	DB 2	Length 270
Best Local Similarity	36.8%	Pred. No. 9, 1e-28		
Matches 100	Conservative 33	Mismatches 121	Indels 18	Gaps 6

  

Qy	49	VAVCVSTPDKQONEKTCALIELVREARLGLACLAFLPEAFDFIARDPAETLHLEPL-GG	107
	11		
Db	6	VAVVMVSGVSEPTVNAAMKRLVARAEAGMVLVLPKXVWLMGANDTKRLALAEPLGGG	65
Qy	108	KLLEETOLARECGIMLSLG-----FHBDGDWEDTQKIYNCHVILNNSKGVAVARIKRT	162
	66		
Db	66	RQTALSETAKCCGYVL-FGTVPLPDSGAG-----KVMNTLLVYGRDGVRTGTYLHKM	117
Qy	163	HLCDVEIPGQPMCESNSTMPGPLESPVSTPAGKTIGLAVCYDMRPEPLSLAQAQAEI	222
	118		
Db	118	HL-FGFSGLGRYAEADTIRAGGDVPHLSAGGVPAAMGICVDYRPE--FERROLPEFV	173
Qy	223	LTFPSAFGSIPTPAHWEVLLRARALETCOCYVAAACCGHHNRASVYGSMDVPMGTAV	282
	174		
Db	174	LMPLPAFTHTTOKAHMELLRARAVDNCQYVAAAGGGLHENGSRKTFGSHMIVDPMGDVL	233
Qy	283	ARCSEGGICLARIDLNYLRQLRHLPLVFOHR	314
	11		
Db	234	DVLPEGEGVATADIDANRLNSVNNRLPALAKYR	265

RESULT 14  
T36488  
probable hydrolase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C:Accession: T36488  
R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.  
Submitted to the EMBL Data Library, July 1999  
A:Reference number: Z21608  
A:Accession: T36488  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-264 <SAU>  
A:Cross-references: EMBL:AL009682; PIDN:CA946930.1; GSPDB:GNO00070; SCQEDB:SCGD3\_08C

A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCODB:SCGD3.08C  
C:Superfamily: hypothetical protein YLR351c

Query Match 19.6%; Score 342.5; DB 2; Length 264;  
Best Local Similarity 35.8%; Pred. No. 7.4e-23;  
Matches 91; Conservative 33; Mismatches 101; Indels 29; Gaps 9;

OY 68 AELVREAAFLGACFLAFLPE-----AFDFTARDPAETLHSEPLGGLLEETYLQALRECL 122  
DB 24 AALVREQA--GADLVVLPFLMTGTAFAFEEDFAA-----AEPRLGPTYEMAKAASDAGV 76  
OY 123 WLSLGFGHERGQDMEOQIKYNCHVLLNSGAVVATRYKTHLCDVEITPGQPMCESNSTM 182  
DB 77 WLHAGSVPRGPD-----GLVNTSLVFSFAGDLTASYRKIHREGPD-----KGAVAL 124  
OY 183 PGPSLESFVST--PAKGIGLAVCYDMRPELSLALQAQAEILTYPSAFGSTITGPAHMEV 240  
DB 125 MGAGRE-PVTVRLPPTTLGATCATCYDLRPELFRSLVDAGAEILVPAWPE--RRRAHWTL 182  
OY 241 LLRAAIEIQCIVVAAAGCGRHHERKASVGHSMVNDPAGTVVARGSEGPLCLARIDLNV 300  
DB 183 LARARAVENQAFLACGTAGTAGVPAQA-GHSIVDPMGEVLAAGADEVLAVEFDPER 241  
OY 301 LRQLRRHLFVFOHR 314  
DB 242 VARTREQFPALKDR 255

RESULT 15  
F75263

probable hydrolase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C:Accession: F75263

R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: F75263

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1 297 <WHI>

A:Cross-references: GB:AE002082; GB:AE000513; NID:95460347; PIDN:AAFL12070.1; PID:9646035

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2535

A:Map position: 1

C:Superfamily: hypothetical protein YLR351c

Query Match 18.2%; Score 318; DB 2; Length 297;  
Best Local Similarity 31.3%; Pred. No. 1.3e-20;  
Matches 92; Conservative 49; Mismatches 117; Indels 36; Gaps 10;

OY 49 VAVCOVSTPRKQONFKTCAELVREARIGACLAFLPEAFD--FIARDPAETLHLESEPL 105  
DB 10 LAVVOMHMTDQLEDVNERAHEVREARAGAOYILLPELFENILYFCQVEREDYFGLAHPL 69  
OY 106 -GGKILLEETOLARECGMLSLGFGHERGQDMEOQIKYNCHVLLNSKGAHVATRYKTHL 164  
DB 70 EGHPTGRFOELARELVNVLPLVSYFEKAG-----QAHNSLVLCIDAGGELLGNRYKTHI 123  
OY 165 CDVEIPGQPMCESNSTM-PGPSLESFVSTPAKGIGLAVCYDMRPELSLALQAQAEIL 223  
DB 124 PD-----GPGYEKRYFPGDGTGRGVGICMOWIFETARVMMLOGADPL 177  
OY 224 TYPSEFGSITGPAH-----WEVLLARAIEIQCIVVAAAGCGRH--HEKRASYGH 271

DB 178 LYPTRIGS--EPAEYETPNNHQOMORAMVGNVSSVYSSNRIGKEIVGLLEQTYGCH 235  
OY 272 SMVNDPMGTVVAR---CSEPGCLARIDLNTLROLRHLFVFOHRRPDLVGNL 322  
DB 236 SFISDTGTGLVALGDSEGP--LLHELNLEARKFRAGMGFFRDRREPLVGPL 287

Search completed: April 26, 2002, 16:55:51  
Job time: 179 sec



**THIS PAGE BLANK (USPTO)**



	DR	MPI: 2000-171195/.15.			
	DR	N-PDSB; AAZ610I..			
	XX	Novel nitricase homologs used as diagnostic and therapeutic reagents	-		
	Pt	for the detection and treatment of cancer -			
	XX				
	PS	Disclosure; Fig 6; 25pp; English.			
	XX				
	CC	The present sequence is encoded by the coding region of human, murine,			
	CC	Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The			
	CC	human and mouse NIT1 genes are members of an uncharacterised			
	CC	mammalian gene family with homology to bacterial and plant nitrilases.			
	CC	The tumour suppressor gene FHIT in D. melanogaster and C. elegans code			
	CC	for fusion proteins in which the Fhit domain is fused with a Nit domain.			
	CC	In mouse and humans, FHIT and NIT are encoded by two different genes,			
	CC	localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The			
	CC	human FHIT gene at chromosome 3p14.2, spanning the constitutive			
	CC	chromosomal fragile site FRAB, is often altered in most common forms			
	CC	of human cancer. The Nit protein overcomes the mutated inactivation			
	CC	of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives			
	CC	and analogues of them, and antibodies are used as diagnostic and			
	CC	therapeutic reagents for the detection and treatment of cancers.			
	SQ				
	XX	Sequence    464 AA;			
	OY	Query Match                  99.7%; Score 1743; DB 21; Length 464;			
		Best Local Similarity 100.0%; Pred. No. 2e-174;			
		Matches 326; Conservative     0; Mismatches      0; Indels       0; Gaps       0			
	Dd	2 LGFTRRPHRPLSLCPGLRIPOLSVCAQPRPRAAISSSCCELPVAVCVSTSPDKD 61			
		gftfrrphrplslcpgrlripqlsvlcqprmaisscceplvavcqvstspdkq  96			
	OY	62 ONFKTCAELVEEARAGACLAFLPEAFDFLARPATTLHSPDGKLEETOLARCGC 121			
	Dd	97 gnfktceelvreaarlgacflapeafdfardpaetlnhselpyggkllseeytlqtarecg 156			
	OY	122 LMTSLGFGHERGDDEGTOKLYNCWHLNSKGAVNATTYRKTHLCDEVIEIPGGPMCESNST 181			
	Dd	157 lwtlsrgfghergddegtoklyncwhlnskgavnattyrtkhcldevieipggpmcesnst 216			
	OY	182 MRGSILSPTVPAGIKGLAVCYDMRPPELLALAAQAEITLTYPASRGSTGPANMEVL 241			
	Dd	217 mrpsilsptvpagikglavcydmrfpellalaaqaellitypasrgstgpamevl   276			
	OY	242 LRARAIFTCGYVAAAOCGRHNHEKRASYGSHMVDPWGVTVARCSEGPSCLARDLNWL 301			
	Dd	277 lraratitgyvaaaqcgrfhckrsyshsmvdpwgvtnarcsegpgscldarlndyl   336			
	OY	302 RQLRRRLPVFOHRHPDLXGNLGHPIS 327			
	Dd	337 rqlrrrlpvfohrhpdlxgnlhghpls 362			
	ID	AAB57054 standard; Protein; 224 AA.			
	XX				
	AC	AAB57054:			
	XX				
	DT	13-MAR-2001 (first entry)			
	DE	Human prostate cancer antigen protein sequence SEQ ID NO:1632.			
	KX				
	KW	Human; prostate cancer; prostate cancer antigen; detection; diagnosis;			
	KW	neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;			
	KW	vulnerary; gastrointestinal; nephrotoxic; antinfecitive; gynaeological;			
	KW	antibacterial; gene therapy; neural; immune; reproductive; renal;			
	KW	gastrointestinal; pulmonary; cardiovascular; proliferative disorder;			
	XZ	wound; infectious disease.			
	OS	Homo sapiens.			

Query Match	Best Local Similarity	53.0%: Score 927; DB 21; Length 224;
Matches 179; Conservative 3; Mismatches 3; Indels 4; Gaps 3		
QY 1 MGGFTTRPHRFLSLCPGLRIPQLSVLCAQDRPRAMAISSSCCLPLVAVCQVSTPDK 60		
DB 26 mlgftrtphrflslcpglripqlsvlcaqdrpramaisssccplvavcqvtatpdk 85		
QY 61 QONFTCAELVBEARLGLACLAFLPAPAFIARDPAETHLSEPLGKLLFEYTOIAREC 120		
DB 86 gqnftlcaelvbeaarlglaclafleaafidlarpaetlhlseplgkllfeetqlarec 145		
QY 121 GLMTSLGGFHEHGOQWEOQKLYNCHVLLNSKGAVAVATYRKTHLCDVEIPGCGPMCESNS 180		
DB 146 glwtslggfhehgoqweoqklynchvllnsgavavatyrtkthlcdvelpgqg-lc-vka 203		
QY 181 TWP--GPST 187		
DB 204 tlpcplgavl 212		
RESULT 3		
ID AAY10877		
AA10877 standard; Protein; 153 AA.		
AA10877;		
AC		
DT 14-MAY-1999 (first entry)		
XX		
XX Amino acid sequence of a human secreted protein.		
XX		
XX Secreted protein; cancer; tumour; neurodegenerative disorder;		
XX		
XX WO200055174-A1.		
XX		
XX 21-SEP-2000.		
XX		
XX 08-MAR-2000; 2000WO-US05988.		
XX		
XX 12-MAR-1999; 99US-0124270.		
XX		
XX (HUMA-) HUMAN GENOME SCI INC.		
XX (ROSE/) ROSEN C A.		
XX		
XX Rosen CA, Ruben SM;		
XX		
XX WPI: 2000-587513/55.		
XX N-PSDB: AAF16257.		
XX		
XX prostate cancer associated gene sequences, referred to as prostate		
XX cancer antigens, useful for treatment, prevention, and diagnosis of		
XX disorders such as prostate cancer -		
XX		
XX Claim 11; Page 2090-2091; 2338pp; English.		
XX		
XX AAF15566 to AAF16505 encode the human prostate cancer associated		
XX proteins, called prostate cancer antigens, given in AAF56363 to AAF57302.		
XX		
XX The prostate cancer antigens can have neuroprotective, cytosolic,		
XX cardioscic, immunomodulatory, muscular, vulnary, gastrointestinal,		
XX nephrotropic, antinefctive, gynaecological and antibacterial activities,		
XX and can be used in gene therapy. The prostate cancer antigen		
XX polynucleotides may be used for detection of prostate cancer, chromosome		
XX identification, as chromosome markers, and for numerous other diagnostic		
XX or research purposes. The prostate cancer antigens may be used to treat		
XX disorders such as neural, immune, muscular, reproductive,		
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative		
XX disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to		
XX AAF57303 represent sequences used in the exemplification of the present		
XX invention.		
XX		
XX Sequence 224 AA;		
XX		

XX developmental abnormality; foetal deficiency; blood disorder;  
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;  
KW renal disease; diabetes; inflammation; allergy; ischemic shock;  
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;  
KW prostate disease; asthma; osteoporosis; arthritis.  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT Peptide 1..46 /note= "signal peptide"  
FT Protein 47..153 /note= "secreted protein"  
XX  
XX WO9907891-A1.  
XX  
XX 18-FEB-1999.  
XX  
XX 04-AUG-1998; 98WO-US16235.  
XX  
XX 19-AUG-1997; 97US-0056732.  
XX 05-AUG-1997; 97US-0054798.  
XX 05-AUG-1997; 97US-0054803.  
XX 05-AUG-1997; 97US-0054804.  
XX 05-AUG-1997; 97US-0054806.  
XX 05-AUG-1997; 97US-0054807.  
XX 05-AUG-1997; 97US-0054808.  
XX 05-AUG-1997; 97US-0054809.  
XX 05-AUG-1997; 97US-0055309.  
XX 05-AUG-1997; 97US-0055310.  
XX 05-AUG-1997; 97US-0055312.  
XX 05-AUG-1997; 97US-0055386.  
XX 05-AUG-1997; 97US-0055311.  
XX 18-AUG-1997; 97US-0055970.  
XX 18-AUG-1997; 97US-0055986.  
XX 19-AUG-1997; 97US-0056365.  
XX 19-AUG-1997; 97US-0056366.  
XX 19-AUG-1997; 97US-0056557.  
XX 19-AUG-1997; 97US-0056370.  
XX 19-AUG-1997; 97US-0056371.  
XX 19-AUG-1997; 97US-0056563.  
XX 19-AUG-1997; 97US-0056731.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;  
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;  
XX  
XX WPI: 1999-167452/14.  
XX N-PSDB: AAX30398.  
XX  
XX New isolated human genes encoding secreted polypeptides - useful for  
FT diagnosis and treatment of pathological diseases  
FT  
XX  
XX Claim 3; Page 304; 331pp; English.  
XX  
XX The specification describes secreted proteins and their corresponding  
XX polynucleotides which are useful for preventing, treating or ameliorating  
XX medical conditions, e.g. by protein or gene therapy. Pathological  
XX conditions can also be diagnosed by determining the amount of the  
XX secreted polypeptides in a sample or by determining the presence of  
XX mutations in the polynucleotides. Specific uses are described for each  
XX of the products, based on which tissues they are most highly  
XX expressed in, and include developing products for the diagnosis or  
XX treatment of cancer, tumours, neurodegenerative disorders, developmental  
XX abnormalities and foetal deficiencies, blood disorders, CNS disorders,  
XX diseases of the immune system, autoimmune diseases, hepatic and renal  
XX diseases, diabetes, inflammation, allergies, ischemic shock, Alzheimer's  
XX and cognitive disorders, schizophrenia, cardiovascular disorders,  
XX prostate diseases, asthma, disorders involving osteoclasts such as  
XX osteoporosis, arthritis or malignancies, diseases of testes, lung or  
XX thymus, digestive/endocrine disorders, infections and AIDS. The  
XX polypeptides are also useful for identifying their binding partners.

XX  
SQ Sequence 153 AA:  
Query Match 47.0%; Score 821; DB 20; Length 153;  
Best Local Similarity 100.0%; Pred. No. 3,3e-78;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 175 KCESNSTMPGSPLESFVSTPAGKIGLAVCYDMRFPPLSLALAOAGELITTPSAFSTTG 234  
DB 1 mcesnstmpgsplesfvsstpagnkiglavcydmrfpelslalaqaealltysafsttg 60  
QY 235 PAHWEVLRLARAIETQCYYVAAAGCGHHKRRASYSCHSNVDPGCVVAARCSGPGICLA 294  
DB 61 panwevllraraletcgyvvaagcgrhnekrrasysghsmvdpwglvvarcsepgpICLA 120  
QY 295 RIDLNTYLRQLRRHLPVFQHRRPDLVGNLGHPLS 327  
DB 121 ridlntylrqlrrhlpvfqhrtrpdlvgnlghpls 153  
RESULT 4  
AAB80984  
ID AAB80984 standard; Protein; 276 AA.  
XX  
XX AAB80984;  
AC  
XX 12-JUN-2001 (first entry)  
DT  
XX  
XX Human hnit3-ase.  
DE  
XX  
XX Human; hnit3-ase; Cushings adrenal gland tumour;  
KW nitrile hydrolytic enzyme; nitrilase.  
XX  
XX Homo sapiens.  
OS  
XX  
XX CN1277998-A.  
PN  
XX  
XX 27-DEC-2000.  
PD  
XX  
XX 30-MAY-2000; 2000CN-0116221.  
PF  
XX  
XX 30-MAY-2000; 2000CN-0116221.  
PR  
XX  
XX (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.  
XX  
XX Xu X, Qian B, Zhang X;  
PI  
XX  
XX WPI: 2001-245678/26.  
XX N-PSDB: AAF84214.  
XX  
XX Human nitrile hydrolytic enzyme protein and its coding sequence -  
PT  
XX  
XX Claim 2; Page 11; 20pp; Chinese.  
PS  
XX  
XX The present invention relates to human hnit3-ase (nitrilase) protein,  
XX which is expressed in Cushings adrenal gland tumour, and its coding  
XX sequence (AAF84214 and AAB80984). The present invention also relates to a  
XX preparation method of said protein and nucleic acid sequence, and a  
XX method of detecting human hnit3-ase nucleic acid sequence and polypeptide  
XX in sample.  
SQ Sequence 276 AA:  
Query Match 27.8%; Score 485.5; DB 22; Length 276;  
Best Local Similarity 36.7%; Pred. No. 1.4e-42;  
Matches 105; Conservative 46; Mismatches 98; Indels 37; Gaps 6;  
QY 49 VAVCQVTSPPDKQONFKCAELVREARLGAACIAFLPEAFDFTARDPAETLHSEPLGSK 108  
DB 6 lallqgssiksdnvttracsflreaatgaktvipecf-----nspygak 52

QY 109 LEEY-----TOLAREGLMLSLGFERGODMEQTORIYNCHVLNSKAVY 156  
 Db 53 yfpeyaekipgestqklsevakceciyligsiipe-----edagkllyntcavfpgdpdgl 107  
 QY 157 AFYRKHLCADVEIRPGCGMCESNSTMPPGSLSPSTAGKIGLAVCYDMRPELSLALA 216  
 Db 108 akyrthlfdldvpgkifgeekltspgdsf-stfdpcrcvylgicymrtlaeqlyya 166  
 QY 217 QAGAEILYPSAFSGITSPAHWEVLLRARAIETOCYVAAACGGRHHEKRASY---GHSM 273  
 Db 157 grgcqlllypgafnlittpahwellqgravdnqyvataaspa-----rdkasyavawghst 223  
 QY 274 VDPKGVTVARCSBPGICLARIDINLYRLRHLDPVFOHRRPDLX 319  
 Db 224 vnpwgevJakagteaalvysdldlklaetrqgipvftrkrsdly 269

RESULT 5  
 AAB42967  
 ID AAB42967 standard; Protein; 159 AA.  
 AC AAB42967;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF2731 polypeptide sequence SEQ ID NO:5462.  
 XX  
 KW Human; open reading frame; ORFX; detection: cytostatic; hepatotropic; anticonvulsant; osteopathic; antiparkinsonian; noctropic; neuroprotective; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatologic; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antineumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkets RA, Leach M;  
 XX  
 DR WPI: 2000-602362/57.  
 DR N-PSDB: AAC77176.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 XX Claim 11; Page 4643; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vlnetary;  
 CC antiparkinsonian; antiparkinsonian; noctropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 159 AA:  
 XX

Query Match 17.7%; Score 310; DB 21; Length 159;  
 Best Local Similarity 46.0%; Pred. No. 1,8e-24;  
 Matches 58; Conservative 24; Mismatches 38; Indels 6; Gaps 2;

QY 197 KIGLAVCYDMRPELSLALQAGAEILYPSAFSGITSPAHWEVLLRARAIETOCYVAA 256  
 Db 30 rvgldgcydmrtlaeqlyagrgcqllypgafnlittpahwellqgravdnqyvata 89  
 QY 257 AOCGRHHEKRASY---GHSMVDPKGVTVARCSBPGICLARIDINLYRLRHLDPVFOH 313  
 Db 90 spa---rdkasyavawghstvnnpwgevJakagteaalvysdldlklaetrqgipvftr 146  
 QY 314 RRPDLX 319  
 Db 147 krsdly 152

RESULT 6  
 AAG92814  
 ID AAG92814 standard; Protein; 266 AA.  
 XX  
 AC AAG92814;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 6568.  
 XX  
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EPI108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (RYOW ) KYOMA HAKKO KOGYO KK.  
 XX  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 DR WPI: 2001-376931/40.  
 DR N-PSDB: AAH68033.  
 XX  
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene







OS Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0138119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151348.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155658.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161358.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 15.4%; Score 270; DB 21; Length 299;

Best Local Similarity 27.6%; Pred. No. 7.3e-20;

Matches 85; Conservative 47; Mismatches 124; Indels 52; Gaps 10;

QY 48 LVAVQVSTPDKOONFTCAELVREARLGLCLAFLEPAF-----DFTARDBA 96  
DB 11 vvsalqfacsdidstnaaerliveaakganilllqellleguyfcaqredffkrakp 70  
QY 97 ETLHLSEPLGLKLEEVYQLARECGILSLGFERGODMEQTKIYNCHVLNKGAVV 156  
DB 71 yknh-----ptlarmkklakeigvnpvsffe-----eantkynsialiddgtdl 117  
QY 157 ATYKTHLCVDEIFPGGPMCESNSTM-DGSLSPSVSTPAGKIGLAVCYDMKFPPLSLAL 215  
DB 118 gilykshpdp-----gpyggekfyfnpdctgfkvfqtkfakigaylciwdgwfpeaararn 171  
QY 216 AQAGAEILLYVSAPGS-----ITGPAHWEVLLRARATETOCYVVAACGCR-----HH 263  
DB 172 vlgaeellfypalqsepdqglidrdhwrtvmqghaanvvlvasnlgkellleteng 231  
QY 264 EKRRAS-YGHSWVVDPMGTIVVARCSE-GPGLCLARIDLNVYLRQLRHLPLVFGHRRPDLV-- 319  
DB 232 psqitfygtsfiagptgeivaeadkseaavlvagfdlamiikrsgwvfyfrrpdljykv 291  
QY 320 -----GNT 322  
DB 292 llmdgnl 299

RESULT 9

AAE9700  
ID AAE9700 standard; Protein; 252 AA.  
XX  
AC AAE9700;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:134.  
XX  
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN W0200100843-A2.  
PD  
XX 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-1B00923.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 09-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032926.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1033004.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 12-AUG-1999; 99US-0148613.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040832.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041380.  
PR 31-AUG-1999; 99DE-1041394.  
PR 31-AUG-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042077.

PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042124.  
PR 03-SEP-1999; 99DE-1042129.  
PR 09-MAR-2000; 2000US-0187970.  
XX  
PA (BAD1 ) BASF AG.  
XX  
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
XX  
XX MPI: 2001-137957/14.  
DR N-PSDB; AAF71819.  
XX  
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
PT pathway proteins, useful for producing fine chemicals in  
PT microorganisms, including organic acids, nonproteinogenic amino acids,  
PT and purine and pyrimidine bases -  
XX  
XX Claim 20; Page 361-362; 1737pp; English.  
PS  
XX  
CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
CC MP nucleic acids are useful for the production of fine chemicals  
CC in microorganisms, including organic acids, nonproteinogenic amino  
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
CC compounds, vitamins, cofactors, polyketides and enzymes.  
XX  
XX Sequence 252 AA:  
SQ  
Query Match 14.7%; Score 256.5; DB 22; Length 252;  
Best Local Similarity 32.5%; Pred. No. 1.5e-18;  
Matches 86; Conservative 36; Mismatches 110; Indels 33; Gaps 10;  
QY 63 NEFTCAELVEARLGLCLFLPEAFPIARDPAETLHL---SEPLGCKLLEETQLARE 119  
DB 3 nfallidaaekaaegarvivfpea---tsgsfgtgrltgqeelgetstavrklade 58  
QY 120 CGLWLSLGFH-----ERGDMEQTKIVNCHVLLNSKGVAVATYRKHLCADVIRPOGR 174  
DB 59 ldvivagmftpadvtvgr--ektlsrvnvtvlisag-ihggnkhiyda-----fg 109  
QY 175 MCSNSTMPGSLSPSTPAGKIGLAVCYDMRPPELSLQAQAEILYPSAFGSITG 234  
DB 110 yresdtkpdel-vvfevddikftvatcydirfpegfdlarnagqllivpitswgd--g 166  
QY 235 PA--HMEVLLRALETQCYVVAQAACGRHNEKR-----ASTGSHMVYDPMGTVVARC 285  
DB 167 pgklegveipraraldstwtivaacgarlpeelrderkptgigshmvtnphgevlasa 226  
QY 286 SEGPGCLARIDLVYLRQLRRHLPV 310  
DB 227 gyepemliadivsglakirealp 251  
RESULT 10  
ID AAB80193 standard; Protein; 252 AA.  
XX  
AC AAB80193;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:1120.  
XX  
XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.  
XX  
XX Corynebacterium glutamicum.  
OS  
XX WO200100843-A2.  
PN  
XX 04-JAN-2001.  
XX  
XX 23-JUN-2000; 2000MO-IB00923.  
PF  
XX  
XX 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 09-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032926.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1033004.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 12-AUG-1999; 99US-0148613.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040832.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041380.  
PR 31-AUG-1999; 99DE-1041384.  
PR 31-AUG-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042124.  
PR 03-SEP-1999; 99DE-1042129.  
PR 09-MAR-2000; 2000US-0187970.  
XX  
XX (BAD1 ) BASF AG.  
XX  
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
PI

DR WPI: 2001-137957/14.  
DR N-PSDB: AAF72312.  
XX  
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
PT pathway proteins, useful for producing fine chemicals in  
PT microorganisms, including organic acids, nonproteinogenic amino acids,  
PT and purine and pyrimidine bases -  
XX  
XX Claim 20; Page 1678-1679; 1737pp; English.  
PS  
XX  
CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
CC pathway (WP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
CC WP nucleic acids are useful for the production of fine chemicals  
CC in microorganisms, including organic acids, nonproteinogenic amino  
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
CC compounds, vitamins, cofactors, polypeptides and enzymes.  
CC  
XX  
SQ Sequence 252 AA:  
  
Query Match 14.7%; Score 256.5; DB 22; Length 252;  
Best Local Similarity 32.5%; Pred. No. 1.5e-18;  
Matches 86; Conservative 36; Mismatches 110; Indels 33; Gaps 10;  
  
QY 63 NFKTCAELVREARLACACLAFLPEAFDLARPATLHL--SEPLGKILEEYQLARE 119  
DB 3 nfallrdaekaaegavylvfpea---tsqsfygrldtqgeeldgefstavrkade 58  
QY 120 CGLMLSLGFFH---ERQDWEQOTKITNCHVNLNSKCAVATYTKTILCYEITRGQP 174  
DB 59 ldvlyvaagmflpadlvqrg---ektlsrvnntvllsqag-lbqgynkllthya----fg 109  
QY 175 MCESNSTMGPSPLESPSPAGKIGLAVCYDMRPFLSLALAOAGAEITLYPSAFGSING 234  
DB 110 yresdvtvkpgdel-vvfeyddikfygatydirfpeqfdlangaqilvwrtsqgd--g 166  
QY 235 PA---HMEVLLRARALETQCYVVAQAQGRHHEKR-----ASYGSMVYVDPEWGVVAVR 285  
DB 167 pklqeqwevlpraratdscwlvacqgarlpeelrderkxprtgigsmvtnpbgvavlaa 226  
QY 286 SEGPGICLARIDLVYRQRLRLPV 310  
DB 227 gyepemladiavsglakirealpv 251  
  
RESULT 11  
AAG09684  
ID AAG09684 standard; Protein; 220 AA.  
XX  
XX AAG09684;  
AC  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7711.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.



PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135639.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147312.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148568.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.





KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
PD 06-SEP-2000.  
XX  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 9905-0121825.  
PR 05-MAR-1999; 9905-0123180.  
PR 09-MAR-1999; 9905-0123548.  
PR 23-MAR-1999; 9905-0125788.  
PR 25-MAR-1999; 9905-0126264.  
PR 29-MAR-1999; 9905-0126785.  
PR 01-APR-1999; 9905-0127462.  
PR 06-APR-1999; 9905-0128234.  
PR 16-APR-1999; 9905-0128714.  
PR 19-APR-1999; 9905-0129845.  
PR 21-APR-1999; 9905-0130077.  
PR 23-APR-1999; 9905-0130449.  
PR 28-APR-1999; 9905-0130510.  
PR 30-APR-1999; 9905-0131449.  
PR 04-MAY-1999; 9905-0132407.  
PR 05-MAY-1999; 9905-0132485.  
PR 06-MAY-1999; 9905-0132486.  
PR 07-MAY-1999; 9905-0132487.  
PR 11-MAY-1999; 9905-0132863.  
PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134221.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139493.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.

PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147132.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.

PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

```

Query Match      12.4%: Score 217.5: DB 21: Length 340;
Best Local Similarity 26.4%: Pred 2.9e-14;
Matches 87: Conservative 47: Mismatches 128; Indels 67: Gaps 14.

QY 41 SSSCEPLVAVCOVYSTFDKQDNFETCAELVEARLACGLFLLEAF----- 88
   ||: : : : : ||: : : : : ||:
Db 16 sstvtvltvgstvyn--dtpatlkhakelfveaaskgkylvlfpfeaflygryprgfrygl 73

QY 89 -----DFIARPAEELHLSF-PLGGKLLLEFYQLARECGSLWLSLGGFHERGDWDETOKI 142
   ||: : : : : ||: : : : : ||:
Db 74 avgyvneegidefrnyhsaalkvppxverlaelagknvlhvmalekdg-----ytl 127

QY 143 YNCNHLILMSKGVAVVYTKRTHLCQVE--IPGGPCMESNSTMPGSPLESFVSPTNPAKGTEL 200
   ||: : : : : ||: : : : : ||:
Db 128 yctalfefspgqfgyghkntkmpstercitcwgg-----dgsitp-----vydattgykisa 177

QY 201 AVCYDMRPEELSLALQAQAEILYTPSAFGSITGPRAHMEVLLRARAITEMOCYVVAANA-C 259
   ||: : : : : ||: : : : : ||:
Db 178 alicwemnpilyrtalakyagiycaptdyals-----ewqasmhlhalevgcfefylsahqfc 233

QY 260 GR-----HHEKASVSGSHVWVDPWGTVVNARC-S-EGPGCLARIDLWY 300
   ||: : : : : ||: : : : : ||:
Db 234 krtefrepnpdyllfindiatkkehdpvtsgvgssvllsplygkvlvlgagpnyesegivltadidlg 293

QY 301 LQGLRRHLVFFQH-RRPDLYGNL---GHP 325
   ||: : : : : ||: : : : : ||:

```

Db	294	iaraklyfdvghyskpdif-nltvnehp	321
RESULT	15		
ID	AA644265		
XX	AA644265 standard; Protein: 346 AA.		
AC	AA644265;		
XX			
DT	18-OCT-2000 (first entry)		
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 55426.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PE	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999; 99US-0121825.		
PR	05-MAR-1999; 99US-0123180.		
PR	09-MAR-1999; 99US-0123548.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0126264.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	06-APR-1999; 99US-0128234.		
PR	08-APR-1999; 99US-0128714.		
PR	16-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	23-APR-1999; 99US-0130891.		
PR	28-APR-1999; 99US-0131449.		
PR	30-APR-1999; 99US-0132048.		
PR	30-APR-1999; 99US-0132407.		
PR	04-MAY-1999; 99US-0132484.		
PR	05-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	06-MAY-1999; 99US-0132487.		
PR	07-MAY-1999; 99US-0132863.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	18-MAY-1999; 99US-0134370.		
PR	18-MAY-1999; 99US-0134768.		
PR	19-MAY-1999; 99US-0134941.		
PR	20-MAY-1999; 99US-0135124.		
PR	21-MAY-1999; 99US-0135351.		
PR	24-MAY-1999; 99US-0135629.		
PR	25-MAY-1999; 99US-0136021.		
PR	27-MAY-1999; 99US-0136392.		
PR	28-MAY-1999; 99US-0136782.		
PR	01-JUN-1999; 99US-0137222.		
PR	03-JUN-1999; 99US-0137528.		
PR	04-JUN-1999; 99US-0137502.		
PR	07-JUN-1999; 99US-0137724.		
PR	08-JUN-1999; 99US-0138094.		
PR	10-JUN-1999; 99US-0138540.		
PR	10-JUN-1999; 99US-0138847.		
PR	14-JUN-1999; 99US-0139119.		
PR	16-JUN-1999; 99US-0139452.		
PR	16-JUN-1999; 99US-0139453.		
PR	17-JUN-1999; 99US-0139492.		
PR	18-JUN-1999; 99US-0139454.		
PR	18-JUN-1999; 99US-0139455.		



```

Db      134 yctalfispqgflgkhrkwmptslercslwgg-----dgstlp-----vydtpigkiga 183
OY      201 AVCYDMRFPPELSIALAQAAGAEILTFPSAFCSITGPAHMEVILARALETQCYVVAQAQ-C 259
Db      184 alcwenmplyrtalyakglelycapladysl---ewqasmihlavegscfvlisahqfc 239
OY      260 GR-----HHEKRASYGSHWVDPMGTFVVARCS-EGPGLCLARIDLNY 300
Db      240 krrelfepndylnfdiivtkkehdpvtvggsvlispigkvlagpnysesglvtadidldgd 299
OY      301 LRQLRRRLPVPQH-RRPDLYGNL---GHP 325
Db      300 iaraklyfdvghyskpdif-nltvnehp 327

```

Search completed: April 26, 2002, 16:54:36  
 Job time: 234 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: April 26, 2002, 16:50:57 ; Search time 28.88 Seconds  
(without alignments)  
254.798 Million cell updates/sec

Title: US-09-357-675C-21

Perfect score: 1748  
Sequence: 1 MLGFTRRPHRFLSLCPGL.....LPVQRRRDLGNLGHPLS 327

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents:AA:\*

1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*\n2: /cgn2\_6/prodata/2/1aa/5H\_COMB.pep:\*\n3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*\n4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*\n5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*\n6: /cgn2\_6/prodata/2/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.5	11.5	303	2	US-08-294-871A-64
2	200.5	11.5	303	3	US-08-876-398A-64
3	199.5	11.4	303	2	US-08-294-871A-26
4	199.5	11.4	303	2	US-08-294-871A-40
5	199.5	11.4	303	2	US-08-294-871A-44
6	199.5	11.4	303	3	US-08-876-398A-26
7	199.5	11.4	303	3	US-08-876-398A-40
8	199.5	11.4	303	3	US-08-876-398A-44
9	198.5	11.4	303	2	US-08-294-871A-6
10	198.5	11.4	303	2	US-08-294-871A-34
11	198.5	11.4	303	2	US-08-294-871A-60
12	198.5	11.4	303	3	US-08-876-398A-6
13	198.5	11.4	303	3	US-08-876-398A-34
14	198.5	11.4	303	3	US-08-876-398A-60
15	197.5	11.3	303	3	US-08-294-871A-62
16	197.5	11.3	303	3	US-08-876-398A-62
17	196.5	11.2	303	2	US-08-294-871A-10
18	196.5	11.2	303	2	US-08-294-871A-12
19	196.5	11.2	303	2	US-08-294-871A-22
20	196.5	11.2	303	2	US-08-294-871A-38
21	196.5	11.2	303	2	US-08-294-871A-42
22	196.5	11.2	303	2	US-08-294-871A-46
23	196.5	11.2	303	2	US-08-294-871A-54
24	196.5	11.2	303	2	US-08-294-871A-58
25	196.5	11.2	303	2	US-08-294-871A-68
26	196.5	11.2	303	3	US-08-876-398A-10
27	196.5	11.2	303	3	US-08-876-398A-12

28	196.5	11.2	303	3	US-08-876-398A-22	Sequence 22, Appl
29	196.5	11.2	303	3	US-08-876-398A-38	Sequence 38, Appl
30	196.5	11.2	303	3	US-08-876-398A-42	Sequence 42, Appl
31	196.5	11.2	303	3	US-08-876-398A-46	Sequence 46, Appl
32	196.5	11.2	303	3	US-08-876-398A-54	Sequence 54, Appl
33	196.5	11.2	303	3	US-08-876-398A-58	Sequence 58, Appl
34	196.5	11.2	303	3	US-08-876-398A-68	Sequence 68, Appl
35	195.5	11.2	303	1	US-07-917-111-2	Sequence 2, Appl
36	195.5	11.2	303	1	US-07-917-111-3	Sequence 3, Appl
37	195.5	11.2	303	1	US-08-479-638-2	Sequence 2, Appl
38	195.5	11.2	303	1	US-08-479-638-3	Sequence 3, Appl
39	195.5	11.2	303	2	US-08-294-871A-2	Sequence 2, Appl
40	195.5	11.2	303	2	US-08-294-871A-4	Sequence 4, Appl
41	195.5	11.2	303	2	US-08-294-871A-8	Sequence 8, Appl
42	195.5	11.2	303	2	US-08-294-871A-14	Sequence 14, Appl
43	195.5	11.2	303	2	US-08-294-871A-16	Sequence 16, Appl
44	195.5	11.2	303	2	US-08-294-871A-18	Sequence 18, Appl
45	195.5	11.2	303	2	US-08-294-871A-20	Sequence 20, Appl

## ALIGNMENTS

RESULT 1  
US-08-294-871A-64  
Sequence 64, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Nanba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiko  
TITLE OF INVENTION: Immobilized Enzyme Preparation and Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990

```

1      FILING DATE: 27-DEC-1990
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: JP 078840/1991
4      FILING DATE: 11-APR-1991
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: JP 140051/1991
7      FILING DATE: 12-JUN-1991
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: PCT/JP92/00739
10     FILING DATE: 10-JUN-1992
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: JP 212692/1992
13     FILING DATE: 10-AUG-1992
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: PCT/JP93/01101
16     FILING DATE: 05-AUG-1993
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: JP 340078/1992
19     FILING DATE: 21-DEC-1992
20     ATTORNEY/AGENT INFORMATION:
21     NAME: Wegner, Harold C.
22     REGISTRATION NUMBER: 25, 258
23     REFERENCE/DOCKET NUMBER: 74129/127/AOPA
24     TELECOMMUNICATION INFORMATION:
25     TELEPHONE: (202) 672-5300
26     TELEFAX: (202) 672-5399
27     TELEX: 904136
28     INFORMATION FOR SEQ ID NO: 64:
29     SEQUENCE CHARACTERISTICS:
30     LENGTH: 303 amino acids
31     TYPE: amino acid
32     TOPOLOGY: linear
33     MOLECULE TYPE: protein
34     OS-08-294-871A-64

```



OY 309 PVF-OHRRDLXG 320  
Db 286 ENFKOHQPHYG 298

## RESULT 3

US-08-294-871A-26

; Sequence 26, Application US/08294871A

; Patent No. 5824522

; GENERAL INFORMATION:

; APPLICANT: Ikenaka, Yasuhiro

; APPLICANT: Namba, Hirokazu

; APPLICANT: Takano, Masayuki

; APPLICANT: Yajima, Kazuyoshi

; APPLICANT: Yamada, Yukio

; APPLICANT: Takahashi, Satomi

; APPLICANT: Okubo, Kazuma

; APPLICANT: Yamada, Kazuhiko

; APPLICANT: Hiraiishi, Yoshiro

; TITLE OF INVENTION: Immobilized Enzyme Preparation and

; TITLE OF INVENTION: Process for Producing D-a-Amino Acid

; NUMBER OF SEQUENCES: 70

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley &amp; Lardner

; STREET: 3000 K Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/294,871A

; FILING DATE: 22-AUG-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/971,758

; FILING DATE: 12-APR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/917,111

; FILING DATE: 07-AUG-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/211,641

; FILING DATE: 11-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 400848/1990

; FILING DATE: 07-DEC-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP91/01696

; FILING DATE: 06-DEC-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 407922/1990

; FILING DATE: 27-DEC-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 078840/1991

; FILING DATE: 11-APR-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 140051/1991

; FILING DATE: 12-JUN-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP92/00739

; FILING DATE: 10-JUN-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 212692/1992

; FILING DATE: 10-AUG-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP93/01101

; FILING DATE: 05-AUG-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP93/01101

; FILING DATE: 05-AUG-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP93/01101

; FILING DATE: 05-AUG-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP93/01101

; FILING DATE: 05-AUG-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 340078/1992

; FILING DATE: 21-DEC-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Wegner, Harold C.

; REGISTRATION NUMBER: 25,258

; REFERENCE/DOCKET NUMBER: 74129/127/AOPA

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 672-5300

; TELEFAX: (202) 672-5399

; TELETYPE: 904136

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 303 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-294-871A-26

## Query Match

Best Local Similarity 11.4%; Score 199.5; DB 2; Length 303;

Matches 83; Conservative 47; Mismatches 129; Indels 49; Gaps 14;

OY 48 LVAVCO---VTSPPDKQNEFKTCAELVREARLGCALFPE--AFDFIAR----DPAET 98  
Db 5 ILAVGQGGPRAARETREQVYVRLDMITRAASGANFIYPELALITTFPRMHTDEAEL 64  
OY 99 LHLSE-PLGKLELEYTOLARECGLWLSIGF--ERGQDWEOTOKIYNCHVLNLSKA 154  
Db 65 DSFEYTEMGPVYRPLPEKAAELGIFNLGYALVYEGG----YKRRTSLIVYKSGK 119  
OY 155 VVATYRKTLCVDEIFGQGMCSNSTMGPSLESPYTP-----AGKIGLAVC 203  
Db 120 IYGYRKHIL-----PGHK---EYEAIRPEQHLKRYFEGDGLGFYVDVDAKMGAFIC 171  
OY 204 YDMRPPELSLALAQAGAEIT--YPSAFGSIQPAHMEVL-----LRARALETCQYV 254  
Db 172 NDRMPPEAMRWMLRGAEIICGYNTPTHNPYQHDHLISPHNLSMGCGSTONGAMSA 231  
OY 255 AAAQCGRHHEKRAISYGHVNDPWGTVVARCSE-GRGLCLARIDLVNLRRLHLPV-Q 312  
Db 232 AAGKVGW-EECNMLLGHSCIVAPTGEIVALTTLLEDEVITAADVADLCRRLRHITFNKQ 290  
OY 313 HRRPDLXG 320  
Db 291 HROPQHYG 298

RESULT 4  
US-08-294-871A-40  
; Sequence 40, Application US/08294871A  
; Patent No. 5824522  
; GENERAL INFORMATION:  
; APPLICANT: Ikenaka, Yasuhiro  
; APPLICANT: Namba, Hirokazu  
; APPLICANT: Takano, Masayuki  
; APPLICANT: Yajima, Kazuyoshi  
; APPLICANT: Yamada, Yukio  
; APPLICANT: Takahashi, Satomi  
; APPLICANT: Okubo, Kazuma  
; APPLICANT: Yamada, Kazuhiko  
; APPLICANT: Hiraiishi, Yoshiro  
; TITLE OF INVENTION: Immobilized Enzyme Preparation and  
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:

```

1 MEDIUM TYPE: floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: Patentin Release #1.0, Version #1.25
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/294,871A
7 FILING DATE: 22-AUG-1994
8 CLASSIFICATION: 435
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 07/971,758
11 FILING DATE: 12-APR-1993
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: US 07/917,111
14 FILING DATE: 07-AUG-1992
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: US 08/211,641
17 FILING DATE: 11-APR-1994
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: JP 400848/1990
20 FILING DATE: 07-DEC-1990
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: PCT/JP91/01696
23 FILING DATE: 06-DEC-1991
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: JP 407922/1990
26 FILING DATE: 27-DEC-1990
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: JP 078840/1991
29 FILING DATE: 11-APR-1991
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: JP 140051/1991
32 FILING DATE: 12-JUN-1991
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: PCT/JP92/00739
35 FILING DATE: 10-JUN-1992
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: JP 212692/1992
38 FILING DATE: 10-AUG-1992
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: PCT/JP93/01101
41 FILING DATE: 05-AUG-1993
42 PRIOR APPLICATION DATA:
43 APPLICATION NUMBER: JP 340078/1992
44 FILING DATE: 21-DEC-1992
45 ATTORNEY/AGENT INFORMATION:
46 NAME: Wegner, Harold C.
47 REGISTRATION NUMBER: 25, 258
48 REFERENCE/DOCKET NUMBER: 74129/127/AOPA
49 TELECOMMUNICATION INFORMATION:
50 TELEPHONE: (202) 672-5300
51 TELEFAX: (202) 672-5399
52 TELEX: 904136
53 INFORMATION FOR SEQ ID NO: 40:
54 SEQUENCE CHARACTERISTICS:
55 LENGTH: 303 amino acids
56 TYPE: amino acid
57 TOPOLOGY: linear
58 MOLECULE TYPE: protein
59
60 US-08-294-871A-40

```

Query Match	11.4%	Score 199.5	DB 2:	length 303:
Best Local Similarity	26.2%	Pred. No. 9.3e-14:		
Matches	82:	Conservative	46:	Mismatches 126: Indels 59: Gaps 14:
OY	48	LVAVQ---	VTSTPKQONFKTCAELVREAAALGCLAFLEP-	AFDFLAR---DPAET 98
		::: ::	::: ::	::: ::
Db	5	ILAVQGGPIAAETREQVAVRLMLTLTAASRGANFIY	PELALTTFFPRMHFTDEAL	64
OY	99	LHLSE-PLGGKLEVEYTOIARECGIMLSIGFRH---	ERQDMEQNKIKINCHYLNSKGA	154
		::: ::	::: ::	::: ::
Db	65	DSFYTEMGPVAVRLFEKAALGIGFNGLVAVELVEGG---	VKRRNTSILVDKSKK	119

OY	155	VAVYRKTHLDVETPGGGPMCESENMPGSLSEPVSTP-----AGTIGTAVC	203
Db	120	IVGKTKRTHL-----PGHK---EYEAATRPQHLERKRFEEBDLGFYYVDAAAMGMC	171
OY	204	YDMRPETSLAAGAEILTPSAFGSITGPAHEVL-----LRARAIET	249
Db	172	NDRMRPEAMRYMGIRGAETIC-----GGYNPTPHNPQHDHLSFHHLLSMQGSYQN	226
OY	250	QCYYVVAALACGRHHEKRASTYGHSMVNDPMGTVVARCSE-GRGLCLATIDANTYLPRRLH	308
Db	227	GAMSAALAKACM-BENCMLLGHSCIVAPTGELVALLTTLLEDEVIITAVDIDRCRELEHI	285
OY	309	PVF-OHRRPDLYG	320
Db	286	FNFQKHQRQPHYG	298

RESULT 5  
 US-08-294-871A-44  
 : Sequence 44, Application US/08294871A  
 : Patent No. 5824522  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Ikenaka, Yasuhiro  
 : APPLICANT: Namba, Hirokazu  
 : APPLICANT: Takano, Masayuki  
 : APPLICANT: Yajima, Kazuyoshi  
 : APPLICANT: Yamada, Yukio  
 : APPLICANT: Takahashi, Satomi  
 : APPLICANT: Okubo, Kazuma  
 : APPLICANT: Yamada, Kazuhiko  
 : APPLICANT: Hiraishi, Yoshiko  
 : TITLE OF INVENTION: Immobilized Enzyme Preparation and  
 : TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
 : NUMBER OF SEQUENCES: 70  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Foley & Lardner  
 : STREET: 3000 K Street, N.W.  
 : CITY: Washington  
 : STATE: D.C.  
 : COUNTRY: U.S.  
 : ZIP: 20007-5109  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentin Release #1.0, Version #1.25  
 :  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/294,871A  
 : FILING DATE: 22-AUG-1994  
 :  
 : CLASSIFICATION: 435  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/971,758  
 : FILING DATE: 12-APR-1993  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/917,111  
 : FILING DATE: 07-AUG-1992  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/211,641  
 : FILING DATE: 11-APR-1994  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: JP 400848/1990  
 : FILING DATE: 07-DEC-1990  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: PCT/Jp91/01696  
 : FILING DATE: 06-DEC-1991  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: JP 407922/1990  
 : FILING DATE: 27-DEC-1990  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: JP 078840/1991  
 : FILING DATE: 11-APR-1991  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: JP 140051/1991

;; FILING DATE: 12-JUN-1991  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/JP92/00739  
;; FILING DATE: 10-JUN-1992  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: JP 212692/1992  
;; FILING DATE: 10-AUG-1992  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/JP93/01101  
;; FILING DATE: 05-AUG-1993  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: JP 340078/1992  
;; FILING DATE: 21-DEC-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Wegner, Harold C.  
;; REGISTRATION NUMBER: 25,258  
;; REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 672-5300  
;; TELEFAX: (202) 672-5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 44:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 303 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-294-871A-44

Query Match 11.4%; Score 199.5; DB 2; Length 303;

Best Local Similarity 26.2%; Pred. No. 9.3e-14; Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 48 LVAVCO---VTSTPDKQONFKTCALVREARLACLAFLPE--AFDFIAR---DPAET 98  
DB 5 ILAVGGGPIARAEETREQVVRLLDMLTKASRGANFVPELALTTFFPRWHTDEAEL 64  
QY 99 LHLSE-PLGGKLEETQALRECGMLSLGFGH---ERGQDWEQOKIYNCHVLLNSKA 154  
DB 65 DSFEYETEMPGVAVRPLFEKKAELGIFNLGYAELVVEGG-----VKRRFNTSLVDKSGK 119  
QY 155 VVATYRKTLCDEVEIPGGPWCESNTPGSLSPVSTP-----AGKIGLAVC 203  
DB 120 IYGYRKIHL-----PGHK---EYEAVRPQHLERKRYEPGDLGFPYVDVDAAMKGMFIC 171  
QY 204 YDMRPFELSLALAOAGAEILTPSAFGSTTGAHWEVL-----LRARAIET 249  
DB 172 NDRMPEARVMGLRGAEILC-----GGYNTPTNPNPLVPOHDLTSFHHLLSMQAGSYON 226  
QY 250 QCYVVAACGGRHHEKRASTGSHMVDPMGTVAARCSE-GPGICLARIDLNYLRRL 308  
DB 227 GAMSAAAGAGM-EENCMILGHSICVAPTEIYALTTLEDEVYTAADVDDRCRELREHI 285  
QY 309 PVF-OHRPDLG 320  
DB 286 FNEKHROPQHTG 298

RESULT 6

US-08-876-398A-26  
Sequence 26, Application US/08876398A

;; PATENT INFORMATION:  
;; APPLICANT: IKENAKA, Yasuhiro  
;; APPLICANT: NANBA, Hirokazu  
;; APPLICANT: TAKANO, Masayuki  
;; APPLICANT: YAJIMA, Kazuyoshi  
;; APPLICANT: YAMADA, Yukio  
;; APPLICANT: TAKAHASHI, Satomi  
;; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
;; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF  
;; NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington

STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876,398A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:

NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399

INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-876-398A-26

Query Match 11.4%; Score 199.5; DB 3; Length 303;

Best Local Similarity 26.9%; Pred. No. 9.3e-14; Matches 83; Conservative 47; Mismatches 129; Indels 49; Gaps 14;

QY 48 LVAVCO---VTSTPDKQONFKTCALVREARLACLAFLPE--AFDFIAR---DPAET 98  
DB 5 ILAVGGGPIARAEETREQVVRLLDMLTKASRGANFVPELALTTFFPRWHTDEAEL 64  
QY 99 LHLSE-PLGGKLEETQALRECGMLSLGFGH---ERGQDWEQOKIYNCHVLLNSKA 154  
DB 65 DSFEYETEMPGVAVRPLFEKKAELGIFNLGYAELVVEGG-----VKRRFNTSLVDKSGK 119  
QY 155 VVATYRKTLCDEVEIPGGPWCESNTPGSLSPVSTP-----AGKIGLAVC 203  
DB 120 IYGYRKIHL-----PGHK---EYEAVRPQHLERKRYEPGDLGFPYVDVDAAMKGMFIC 171  
QY 204 YDMRPFELSLALAOAGAEILT--YPSAFGSTTGAHWEVL-----LRARAIETOCYV 254  
DB 172 NDRMPEARVMGLRGAEILICGGYNTPTNPNPLVPOHDLTSFHHLLSMQAGSYONGAMSA 231  
QY 255 AAACGGRHHEKRASTGSHMVDPMGTVAARCSE-GPGICLARIDLNYLRRLPVF-O 312  
DB 232 AAGVVG-M-EENCMILGHSICVAPTEIYALTTLEDEVYTAADVDDRCRELREHI 290  
QY 313 HRPDLG 320  
DB 291 HROPQHTG 298



Best Local Similarity 26.28; Pred. No. 9.3e-14;  
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

```

OY 48 IAWCQ---VSTPDKOONFTECAEILYBEARAGCACLAPPE--AFDFIAR---DPAET 98
D 5 ILANGQGFIAKETREBYVVRLLDMILTKAASRKANITVEPALLTTFPRRWFTTDAEL 64
OY 99 LHLSF-PLGKGLBEYTOLARCEGLMLTSLGFFH--ERGODWEOFTOKIYNCHVLLNSGA 154
D 65 DSFEYETEMRPGVVRPLPEFKAAELIGFNLAGVELVVBEGG-----YKRRFNTSILYVDSGK 119
OY 155 VVATYRKTHLCDVEIPGGGPMCESNTPMGPSLESVPSTP-----AGKIGLAVC 203
D 120 IVGKYRKTHL-----PGIK---EYEAIRRPQHLKRRFFEEODLGFVYVDDAKMGHFC 171
OY 204 YDMRPELISLALAOAGAILTYPSAFSITGPAHVEYL-----LRARIET 249
D 172 NDRMRPEMVRMVGRLGAEIILC---GGYPTFNPNPLVPOHDILTSFHNLLSMOAGSYON 226
OY 250 QCYUVAALACRHHHEKRSYSHVMVDPMGVAVRCSF-GRGLCLARIDILNYLQOLRHL 308
D 227 GAMSAAAGKAKM-BENCMLLGHSCIVAVPTGIEVALTTTLEDEYVTAAVDDCRKRELRHI 265
OY 309 PVF-QHRRPDLTG 320
D 286 FNFQOHRQPOHYG 298

```

RESULT 9  
 US-08-294-871A-6  
 Sequence 6, Application US/08294871A  
 Patent No. 5824522  
 GENERAL INFORMATION:  
 APPLICANT: Ikenaka, Yasuhiro  
 APPLICANT: Namba, Hirokazu  
 APPLICANT: Takano, Masayuki  
 APPLICANT: Yajima, Kazuyoshi  
 APPLICANT: Yamada, Yukio  
 APPLICANT: Takahashi, Satomi  
 APPLICANT: Okubo, Kazuma  
 APPLICANT: Yamada, Kazuhiko  
 APPLICANT: Hiraiishi, Yoshiko  
 TITLE OF INVENTION: Immobilized Enzyme Preparation and  
 Process for Producing D-Amino Acids  
 NUMBER OF SEQUENCES: 70  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/294,871A  
 FILING DATE: 22-AUG-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/971,758  
 FILING DATE: 12-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/917,111  
 FILING DATE: 07-AUG-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/211,641  
 FILING DATE: 11-APR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 400848/1990  
 FILING DATE: 07-DEC-1990

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER: PCT/JP91/01696
3      FILING DATE: 06-DEC-1991
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: JP 407922/1990
6      FILING DATE: 27-DEC-1990
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: JP 078840/1991
9      FILING DATE: 11-APR-1991
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER: JP 140051/1991
12     FILING DATE: 12-JUN-1991
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER: PCT/JP92/00739
15     FILING DATE: 10-JUN-1992
16     PRIOR APPLICATION DATA:
17     APPLICATION NUMBER: JP 212692/1992
18     FILING DATE: 10-AUG-1992
19     PRIOR APPLICATION DATA:
20     APPLICATION NUMBER: PCT/JP93/01101
21     FILING DATE: 05-AUG-1993
22     PRIOR APPLICATION DATA:
23     APPLICATION NUMBER: JP 340078/1992
24     FILING DATE: 21-DEC-1992
25     ATTORNEY/AGENT INFORMATION:
26     NAME: Wegner, Harold C.
27     REGISTRATION NUMBER: 25,258
28     REFERENCE/DOCKET NUMBER: 74129/127/AOPA
29     TELECOMMUNICATION INFORMATION:
30     TELEPHONE: (202) 672-5300
31     TELEFAX: (202) 672-5399
32     TELEX: 904136
33     INFORMATION FOR SEQ ID NO: 6:
34     SEQUENCE CHARACTERISTICS:
35     LENGTH: 303 amino acids
36     TYPE: amino acid
37     TOPOLOGY: linear
38     MOLECULE TYPE: protein
39     US-08-294-871A-6

```

[illegible]

Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Namba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraiishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399

TELEX: 904136  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-294-871A-34  
Query Match 11.4%, Score 198.5, DB 2, Length 303,  
Best Local Similarity 26.2%, Pred. No. 1.2e-13,  
Matches 82, Conservative 46, Mismatches 126, Indels 59, Gaps 14,  
OY 48 LVAVCQ---VTSTPKQKQNFYKTCALVREARLGAFLPE--AFDFIAR---DPAET 98  
DB 5 ILAVGQGPRIAGAEETREQVYVRLDMLTKAASRGANFIYFPLALITFFPRRYFIDEAL 64  
OY 99 LHLSE-PLGKLLIEYTLARECGMLSLGFFH---ERGQDMEQTKIYNCHVLLNSGA 154  
DB 65 DSFYETEMPGPVVRPLFEKAELGIGFNLGYALVEGG-----VKRRFNTSILVDSGK 119  
OY 155 VVATYRKTHLCVLEIPGCPMCESTNTPGSLSPVSTP-----AGKIGLANVC 203  
DB 120 IYKRYKRIHL-----PGHK---EYDAVYRPOHLERKYEPPGDLGFPYVDVDAKMGMTIC 171  
OY 204 YDMRPELSIALAQCAELITYPASAFSGITGPAHNEVL-----LARALET 249  
DB 172 NDRWRPEAMRVVGLGALGELIC-----GGYNTPTNPIVQHDHLTSFHHLLSMQAGSYCN 226  
OY 250 QCYVVAACGGRHNEKRASYSYGVNDPWGVVANCSE-GPGLCLARIDLNTLRRL 308  
DB 227 GAMSAAAGKVGW-ENECMLHGSHCIAPGCIYALTTLEDEVITAAVDLDCRELEH 285  
OY 309 PVF-QHRRPDLTG 320  
DB 286 FNFQKROPQHYG 298  
RESULT 11  
US-08-294-871A-60  
Sequence 60, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Namba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraiishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

```

1 APPLICATION NUMBER: US 07/971,758
2 FILING DATE: 12-APR-1993
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 07/917,111
5 FILING DATE: 07-AUG-1992
6 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: US 08/211,641
8 FILING DATE: 11-APR-1994
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: JP 400848/1990
11 FILING DATE: 07-DEC-1990
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: PCT/JP91/01696
14 FILING DATE: 06-DEC-1991
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: JP 407922/1990
17 FILING DATE: 27-DEC-1990
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: JP 078840/1991
20 FILING DATE: 11-APR-1991
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: JP 140051/1991
23 FILING DATE: 12-JUN-1991
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: PCT/JP92/00739
26 FILING DATE: 10-JUN-1992
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: JP 212692/1992
29 FILING DATE: 10-AUG-1992
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: PCT/JP93/01101
32 FILING DATE: 05-AUG-1993
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: JP 340078/1992
35 FILING DATE: 21-DEC-1992
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Wegner, Harold C.
38 REGISTRATION NUMBER: 25,258
39 REFERENCE/DOCKET NUMBER: 74129/127/AOPA
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: (202) 672-5300
42 TELEFAX: (202) 672-5399
43 TELEX: 904136
44 INFORMATION FOR SEQ ID NO: 60:
45 SEQUENCE CHARACTERISTICS:
46 LENGTH: 303 amino acids
47 TYPE: amino acid
48 TOPOLOGY: linear
49 MOLECULE TYPE: protein
50
51 US-08-294-871A-60

```

```

Query Match      11.4%: Score 198.5: DB 2: Length 303:
Best Local Similarity 26.2%: Pred. No. 1.2e-13:
Matches 82: Conservative 46: Mismatches 120: Indels 59: Gaps 14:

OY 48 LVACQ---VSTPDKQONFKTCAELVREARLACLAFLPE--AFDIAR---DPAET 98
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 ILVAGQGPPIRAETREQVVVRLMLTLTKASRGANFIVPELALTTFFPRVYFTDEAL 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 99 LHLSE-PLGGKLEBYTOLARECGLMLSLGGPH---ERQDMEQOKITNCVHLNSKA 154
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 DSFYTEMGPVYVRLFEAAELGIGFNIGYELVEVG---YKRPNISILVDSCK 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 155 VVAIYRKTHLDELPGGPMCESNSTMPGSLSPVSTP-----AGKTGLAVC 203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 IVGKRRKTHL-----PGHK---EYAYRPFQHLERKRYEPGDLGPVYDVDAKKMGMFIC 171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 204 YDMRPEPISIALAQAAGELTTPSAFGSITGAHMEVL-----LRARLET 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 NDRMRPEAKRWGLGAELIC-----GGYNTPTNPLVPHOHLTSFHHLSLMGOSTYN 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 250 QCYVVAQAACGGHHHEKRASYGSHSNVVDWPGTVVAVRCSE-GPGICLARIDILNROLRRHL 308

```

Dd 227 GAMSAAKSSM-BENCMLHSHCIVAPTEGIALTTTLEDEVITALAVIDDRCRELPHEI 285  
QY 309 PVF-OHRRPDLYG 320  
Db 286 FNFKOHRQPCHYG 298

```

: RESULT 12
: US-08-876-398A-6
: Sequence 6, Application US/08876398A
: Patent No. 6083752
: GENERAL INFORMATION:
:   APPLICANT: IKENAKA, Yasuhiro
:   APPLICANT: NAKABA, Hirokazu
:   APPLICANT: TAKANO, Masayuki
:   APPLICANT: YAJIMA, Kazuyoshi
:   APPLICANT: YAMADA, Yukio
:   APPLICANT: TAKAHASHI, Satomi
:   TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
:   TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
:   NUMBER OF SEQUENCES: 70
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: FOLEY & LARDNER
:   STREET: 3000 K Street, N.W.
:   CITY: Washington
:   STATE: D.C.
:   COUNTRY: U.S.A.
:   ZIP: 20007-5109
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentln Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/876,398A
:   FILING DATE: 16-JUN-1997
:   CLASSIFICATION: 435
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US 08/211,641
:   FILING DATE: 11-APR-1994
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: WO PCT/JP93/01101
:   FILING DATE: 05-AUG-1993
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: JP 340078/1992
:   FILING DATE: 21-DEC-1992
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: JP 213692/1992
:   FILING DATE: 10-AUG-1992
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Wegner, Harold C.
:   REGISTRATION NUMBER: 25,258
:   REFERENCE/DOCKET NUMBER: 74129/130
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (202) 672-5300
:   TELEFAX: (202) 672-5399
:   INFORMATION FOR SEQ ID NO: 6:
:   SEQUENCE CHARACTERISTICS:
:   LENGTH: 303 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
:   MOLECULE TYPE: protein
:   US-08-876-398A-6

```

	Query Match	Similarity	26.2%	Pred	198.5	DB	3;	Length	303;
db	Match	Local Similarity	82;	Conservative	46;	Mismatches	126;	Indels	59;
Oy	48	LVAACQ---	VTPSPKQONFKTCAELVREARLGLACLAFLPE--AFDPTAR---	DPAET	98				
		:::::	:::::	:::::	:::::				
	5	ILAAQGGPIARAAREGQVVVLRLDLMLTKASRGNGNFIYFPRLLATLTFFRRHNFDEAL	64						





```

1      APPLICATION NUMBER:  JP 340078/1992
2      FILING DATE:  21-DEC-1992
3      PRIOR APPLICATION DATA:
4
5      APPLICATION NUMBER:  JP 212692/1992
6      FILING DATE:  10-AUG-1992
7
8      ATTORNEY/AGENT INFORMATION:
9
10     NAME:  Wegner, Harold C.
11     REGISTRATION NUMBER:  25,258
12     REFERENCE/DOCKET NUMBER:  74129/130
13
14     TELECOMMUNICATION INFORMATION:
15
16     TELEPHONE:  (202) 672-5300
17     TELEFAX:  (202) 672-5399
18
19     INFORMATION FOR SEQ ID NO: 60:
20
21     SEQUENCE CHARACTERISTICS:
22
23     LENGTH:  303 amino acids
24     TYPE:  amino acid
25     TOPOLOGY:  linear
26
27     MOLECULE TYPE:  protein
28
29     ;
30     ;
31     US-08-876-398A-60

```

```

Db 65 DSFYETEMPGPVVRPLFEKAEIIGFNLGVALVEEG-----VKRRNTSILYDKSGK 119
Oy 155 VVATYRKTHLCOVETIPGOGPMCESNSTMPGSPLESVPSTP-----AGKTGLAVC 203
Db 120 IVGKTRKTHL-----PGHK---EYEAIRPFQHLERKRYEFGDLGFPVYDVDAKMGMPIC 171
Oy 204 YDMRPEELSLALQAQAEILTYPSAFGSITGPAHWEVL-----LRARALET 249
Db 172 NDRMPDEARVMVGLRGAELIC-----GGYNTPTHNPAPVQHDHLTSFHHLSMQAGSYCN 226
Oy 250 OCYVVAACCGRHHEKRASYGSHWVDPWGTIVVARCSE-GPGICLARIDNLNLRQLRHL 308
Db 227 GAMSAAAGVGM-ENCMMLGHSCIYAPRGIEIVALTTLEDEVITAAVDLDRCRELREHI 285
Oy 309 PVF-OHRRPDLG 320
Db 286 FNFKQHRQPOHYG 298

```

Search completed: April 26, 2002, 16:55:11  
 Job time: 254 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 17:02:49 ; Search time 21.41 Seconds

(without alignments)  
553.141 Million cell updates/sec

Title: US-09-357-675c-22

Perfect score: 1736

Sequence: 1 MGFTRPPHQLCTGYRL.....LPVQRHRRPDLVGSIGHPLS 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	523	30.1	307	1 YJM6_YEAST	P47016 saccharomyc
2	492	28.3	272	1 Y601_SYNY3	P55175 synchocyst
3	437.5	25.2	322	1 YAUB_SCHPO	Q10166 schizosacch
4	419	24.1	291	1 YL85_YEAST	P49954 saccharomyc
5	364	21.0	262	1 YBEM_ECOLI	P39874 escherichia
6	361	20.8	262	1 YBEM_ECO57	P58054 escherichia
7	268.5	15.5	340	1 Y480_MYCTU	Q11146 mycobacteri
8	268	15.4	285	1 YPOQ_PSEFL	P55176 pseudomonas
9	249.5	14.4	261	1 YAG5_STRAU	P55177 staphylococ
10	237.5	13.7	234	1 YAG5_STRAU	P55178 staphylococ
11	206	11.9	346	1 NR1L_ARATH	P32861 arabidopsi
12	203	11.7	346	1 NR1L_ARATH	P46010 arabidopsi
13	193	11.1	393	1 BOP_RAT	Q03248 rattus norv
14	187.5	10.8	339	1 NR1L_ARATH	P32862 arabidopsi
15	186.5	10.7	346	1 ALAM_PSEAE	P11436 pseudomonas
16	181.5	10.5	513	1 YHXC_BACSU	P54608 bacillus su
17	170	9.8	355	1 NR1L_ARATH	P46011 arabidopsi
18	168.5	9.7	345	1 ALAM_RHOER	Q01360 rhodococcus
19	156	9.0	349	1 NR1L_TOBAC	Q42965 nicotiana t
20	143	8.5	256	1 YAEV_ECOLI	Q47679 escherichia
21	143	8.2	256	1 NAER2_THEMA	Q9X0Y0 thermotoga
22	126	7.3	356	1 NR1L_ALCPA	P20960 alcaligenes
23	123.5	7.1	567	1 MADE_ADUAE	O67091 aquifex aeo
24	117.5	6.8	368	1 CYH4_GLOSO	P32964 glaucoceros
25	112.5	6.5	199	1 Y1Q4_YEAST	P40447 saccharomyc
26	111.5	6.4	552	1 MADE_RHOCA	Q03638 rhodobacter
27	108	6.2	519	1 LNT_SYNY3	P74055 synchocyst
28	107	6.2	383	1 NR1L_RHORI	Q02068 rhodococcus
29	106.5	6.1	679	1 MADE_MYCTU	P71111 mycobacteri
30	100	5.8	794	1 SEIL_HUMAN	Q929V2 homo sapien
31	99	5.7	541	1 LNT_CHLUP	Q92791 chlamydia p
32	98.5	5.7	511	1 LNT_PSEAE	Q92186 pseudomonas
33	97.5	5.6	365	1 NR1L_RHORI	Q03217 rhodococcus

34	96.5	5.6	680	1 MADE_MYCLE	Q9cbz6 mycobacteri
35	95	5.5	512	1 VNN1_MOUSE	Q920t8 mus musculu
36	93.5	5.4	455	1 TMS5_MOUSE	O9er04 mus musculu
37	93	5.4	425	1 YGEP_ECOLI	P520t8 escherichia
38	92.5	5.3	542	1 LNT_CHLUR	O84539 chlamydia t
39	92	5.3	790	1 SEIL_MOUSE	Q92296 mus musculu
40	91.5	5.3	513	1 VNN1_HUMAN	O95497 homo sapien
41	91.5	5.3	520	1 VNN2_HUMAN	O95498 homo sapien
42	90.5	5.2	541	1 LNT_RHIME	O52910 rhizobium m
43	90	5.2	349	1 NR1L_KLEPO	P100t5 klebsiella
44	87.5	5.0	496	1 LNT_RICPR	O92dg3 rickettsia
45	86.5	5.0	500	1 VNN3_MOUSE	Q9qz25 mus musculu

## ALIGNMENTS

RESULT	ID	YJM6_YEAST	STANDARD	PRT	307 AA.
AC	P47016				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	HYPOHETICAL 34.7 KDA PROTEIN IN SPT10-GCD14 INTERGENIC REGION.				
GN	YJL126W OR J0706.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_Taxid=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRATIN=5288C / FY1679;				
RA	MDLINE=97103775; PubMed=8948101;				
RA	C2iepluch C., Kordes E., Pujol A., Jauniaux J.-C.;				
RT	"Sequencing analysis of a 40.2 kb fragment of yeast chromosome X				
RT	reveals 19 open reading frames including URA2 (5' end), TRX1, PBS2,				
RT	SPT10, GCD14, RPE1, PHO86, NCA3, ASR1, CCR7, GZF3, two tRNA genes,				
RT	three remnant delta elements and a Ty4 transposon."				
RL	Yeast 12:1471-1474(1996).				
CC	- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL: Z49401; CAA89421.1; -				
DR	SGD: S0003662; YJL126W.				
DR	InterPro: IPR003010; CN_hydrolase.				
DR	InterPro: IPR001110; UPF0012.				
DR	Pfam: PF00795; CN_hydrolase; 1.				
DR	PROSITE: PS01227; UPF0012; 1.				
KW	Hypothetical protein.				
SO	SEQUENCE 307 AA; 34693 MW; 48787C43B10A828E CRC64;				
Query Match	30.1%; Score 523; DB 1; Length 307;				
Best Local Similarity	38.3%; Pred. No. 5e+40;				
Matches 116; Conservative 55; Mismatches 104; Indels 28; Gaps 6;					
QY	41 ELPLVAVCOVTSFENKQENFTCAELVQEARLQACIAFLPEAFDFTARNPAETILL--				97
DB	4 LKRVAVAAQGLSSADLTKNLKLVKELISEAQKADAVFLPEASDYISQNLFSRYLAOK 63				
QY	98 SEPLNGLLCOYSOLARE--CGIWLSLG-GFHEGQD-WEDNQKITYNCVILNKGSYVA 153				
DB	64 SPKFIROILOSITDVLVDNSRNIDVSIQVHLPPSEODLLEGNDRVNVLLYIDHREGTILQ 123				
QY	154 SYRTHICDVEIIPGQGMRESNYYTKPGGTLEPPYKTAGKGLAICVDMRPEELSLKLAQ 213				

ID	Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
Db 124	EVQKLTLEFDVDPNPGILAKESKSPGKAIPIILIESPLGKLSACIYDIRPFESLKLRS	41.6%	38	106	16	5	
QY 214	ACAELLITPVSAGSVGTGAHMEVLLRARIESCQYVIAAOCGRH-----	28.3%	492	DB 1	Length 272		
Db 184	MCCELLCPESAFITKTEGEHMELLGRARAVDPQCVCYVLMGQVGMHDLSDPEWOKOSHMSA	41.6%	38	106	16	5	
QY 259	---HEFRASVGSMSVDDPVCATVANC---SESPGICLARIIDLHFLQOMRQHLPYVOHRRP	28.3%	492	DB 1	Length 272		
Db 244	LEKSSRRSSGMSVADIPWGKIIAHADPSTVGPOLLADLDRELLQETRNKMPLMNRRD	41.6%	38	106	16	5	
QY 313	DLY 315	28.3%	492	DB 1	Length 272		
Db 304	DLY 306	41.6%	38	106	16	5	
RESULT 2							
ID	Y601	SYNY3	STANDARD	PRF	272 AA		
AC	P55175						
DT	01-OCT-1996	(Rel. 34, Created)					
DT	01-OCT-1996	(Rel. 34, Last sequence update)					
DT	20-AUG-2001	(Rel. 40, Last annotation update)					
DE	HYPOTHETICAL 30.2 KDA PROTEIN SLD0601.						
GN	SLD0601.						
OS	Synechocystis sp. (strain PCC 6803).						
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.						
OX	NCBI_Taxid=1148;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=96127529; PubMed=8590279;						
RA	Kaneko T., Tanaka A., Sato S., Kocani H., Sazuka T., Miyajima N.,						
RA	Sugita M., Tabata S.,						
RT	"Sequence analysis of the genome of the unicellular cyanobacterium						
RT	Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb						
RL	region from map positions 64% to 92% of the genome.";						
RL	DNA Res. 2:153-166(1995).						
CC	-I- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.						
CC	-----						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -						
CC	the European Bioinformatics Institute. There are no restrictions on its						
CC	use by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>						
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).						
CC	-----						
DR	EMBL: D64002; BAA10370.1; -						
DR	InterPro: IPR003010; CN_hydrolyase.						
DR	InterPro: IPR001110; UPF0012.						
DR	Pfam: PF00795; CN_hydrolyase; 1.						
DR	PROSITE: PS01227; UPF0012; 1.						
KW	Hypothetical protein; Complete proteome.						
SEQ	SEQUENCE 272 AA: 30191 MW; 7787BA9B2BE003A1 CRC64;						

Dd	176	DVLFPAPFAFTATGCDHNOVLLQAAIAIENTCYVIAPATGTCHYERHRHHGHAMIIDPAGV	235
Oy	277	VVARCEGGPLCLARIDLHFLOQRHQLPVROHR	310
Dd	236	ILADAGEKGLAIEINPDRLKOVROOMPISLOHR	269
RESULT 3			
ID	YAUB_SCHPO	STANDARD:	PRT: 322 AA.
AC	Q10166;		
Dt	01-OCT-1996 (Rel. 34, Created)		
Dt	01-OCT-1996 (Rel. 34, Last sequence update)		
Dt	01-NOV-1997 (Rel. 35, Last annotation update)		
Dt	HYPOTHETICAL 35.7 KDA PROTEIN C26A3.11 IN CHROMOSOME I.		
DE	SPAC26A3.11.		
OS	Schizosaccharomyces pombe (fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
CC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OX	NCBI_Taxid=4896;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-972;		
RL	McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V. ;		
RA	Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.		
CC	-1- SIMILARITY: BELONGS TO THE UPE0012 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL; 269240; CAAG3234.1; -		
DR	InterPro: IPR003010; CN_hydrolase.		
DR	InterPro: IPR001110; UPE0012.		
DR	Pfam: PF00795; CN_hydrolase; 1.		
KW	PROSITE: PS01227; UPE0012; 1.		
DR	Hypothetical protein.		
SO	SEQUENCE 322 AA; 35679 MW; EA6F39B160C7F49F CRC64;		
Query Match 25.2%; Score 437.5; DB 1; Length 322;			
Best Local Similarity 33.4%; Pred. No. 2.9e-32;			
Matches 102; Conservative 51; Mismatches 113; Indels 39; Gaps 5;			
Oy	34	MSSSTSWELP-----IIVAVCOVTSTPNKOENFKTCAELVOEABRAGLAFLPEAFDFI	87
Dd	29	MSVSASSLVPPDGFRAFRIGLVOLANTTKDKSEMLALARKLYVEAKANGSNVITLPEIF--	85
Oy	88	ARNPAETILLISEPLNGDLGOY-----SQLARECGIWLSLSGGFHEHGOME	133
Dd	86	-----NSPYGTGYENOYAEPIESSPSYQALSMMADMDTKYILYGGSIPER-----	130
Oy	134	ONOKIYNCHNLNSGVSVAIRKTHLDVELPGGGRRESNYTRPGGLEPPVTPAGK	193
Dd	131	KDGKIYNPAMVPDSGKLIAVRKHLDLIDIPGVSVRESUSLSPGAM-TMWDLEYGK	189
Oy	194	VGLAICTYMRPELSLKLAGAIELTFPSAFSGVSTGAHNEVLLRAAISOCVIVAA	253
Dd	190	FGLIGCIYDIRPEPLAMIARAAGCSMIITPGAFLNSTGPLMHELLARARAVDNEMFACCA	249
Oy	254	QCGRHHTFRASYGSMVVDPMGTVVAVARCSEGGCLCLARIDLHFLQQRHQLPVROHRPD	313
Dd	250	PARDNMADYHSWGSHSTVVDPEGKVATIDEKPSIYAYADIDSVMSTANSVPITYTORRD	309
Oy	314	LYGSL	318
Dd	310	YSEV	314

```

RESULT 4
YL85_YEAST STANDARD; PRT; 291 AA.
ID YL85_YEAST
AC P49554;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOHETICAL 32.5 KDA PROTEIN YLR351C.
GN YLR351C OR L9638.5.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Knudsen T., Hallsworth K., Hawkins J., Hallier L., Jier M.,
RA Johnson D., Johnston L., Langston J., Latreille P., Le T.,
RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Ritken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U19102; AAB67751.1;
DR SGD: S0004343; YLR351C.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
DR KW Hypothetical protein.
SQ
SEQUENCE 291 AA; 32549 MW; AB13744120088827 CRC64;

```

```

Query Match 24.1%; Score 419; DB 1; Length 291;
Best Local Similarity 33.8%; Pred.No.1.2e-30;
Matches 99; Conservative 48; Mismatches 110; Indels 36; Gaps 6;

```

```

QY 45 VAVCVQVT-STPKQKNFTCAELVQEAR--LGACIAFPPEAFDFIARPAETLLISEPL 101
DB 13 VALVOLSSSSPDKMANIOLRAATFIERAKKEQPDTRLVLPPECF-----NSBY 59
QY 102 NGDLLGOYSQ-----LAEEGIMTSLGFFHERGQDWEONOKIYNCHVL 145
DB 60 STDQPKRSEVINKEPSTSVQVFLSNLANKKEIILVGGTIPELD---PRTKIYNTSIIF 116
QY 146 NKSQSVASVYRKLCDVEIPQGGPMRBSNYTRPGCTLEPPVKTAGKGLAICYDMRP 205
DB 117 NEDGLIDKHRRVHLFDVIDPIMGISFHESETLSP-GEKSTIDTKYKRGVGYCDMRP 175
QY 206 ELSLKLAAGAEILTPPAFGSVTPPAHWEVLLRARAISQCYVIAAOCGHHETRAST 265
DB 176 ELAMLSAKKGAFAAMITYPAPFNTVGTPLHWHLLARSAVDNOYVIMCSPARNLQSSYAY 235
QY 266 GHSWVVDWGVAVARCSGPGICLARIIDLHFLQOKRQHLVPVORHRRPDYGSU 318
DB 236 GHSIVVDPRKGIIVALEGGEELITVADLPDEVLESFRQAVPLTKQRFDYISDV 288

```

```

RESULT 5
YBEM_ECOLI

```

```

ID YBEM_ECOLI STANDARD; PRT; 262 AA.
AC P39874; P77192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL PROTEIN YBEM.
GN YBEM OR B0625/B0626.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RL
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alpa H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Samei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RT DNA Res. 3:137-155(1996).
RL
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Kemp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-51 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95075315; PubMed=7984109;
RA Yamanaka K., Milani T., Ogura T., Niki H., Hiraga S.;
RT "Cloning, sequencing, and characterization of multicopy suppressors
RT of a mukB mutation in Escherichia coli."
RL Mol. Microbiol. 13:301-312(1994).
RN [5]
RP SEQUENCE OF 31-78 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Robison K., O'Keefe T., Church G.M.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [6]
RP CONCEPTUAL TRANSLATION.
RA Budd K.E.;
RL Unpublished observations (JAN-2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE K12 SEQUENCE
CC DIFFERS FROM THE WILD-TYPE SEQUENCE DUE TO A FRAMESHIFT IN
CC POSITION 67.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000167; AAC73726.1; ALT_FRAME.
CC DR AEO00167; AAC73727.1; ALT_FRAME.
CC EMBL: D90703; BAA35268.1; ALT_FRAME.

```



RC STRAIN-CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickley E.,  
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolova M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,  
 RA Bishal W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: Z77162; CAB00941.1; -  
 DR EMBL: AF006951; AAK44721.1; -  
 DR TIGR: MT0498; -  
 DR TubercuList; RV0480c; -  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR001110; UPF0012.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS01227; UPF0012; 1.  
 DR Hypothetical protein; Complete proteome.  
 KM Hypothetical protein; VAL-RICH.  
 FT DOMAIN 192 201  
 SQ SEQUENCE 340 AA; 35195 MW; 541EAC9CD1D439B9 CRC64;

Query Match 15.5%; Score 268.5; DB 1; Length 340;  
 Best Local Similarity 28.6%; Pred. No. 6.1e-17;  
 Matches 94; Conservative 41; Mismatches 133; Indels 61; Gaps 11;

QY 18 RLRLRPV-----LCTQPRPRTSSSTSWELPLVAVCQVSTPNK 56  
 |||||  
 DB 24 RLRLRPVRRARRRAOAGLPSCARRGALVAGPRLARR-----IALQIISGDP 74  
 QY 57 QENFKTCAELVOEARILGACLAFLPEAFDFTARNPAETLLSEPLNGDLLGOYSOLAREC 116  
 |||||  
 DB 75 AANQLVKGAYGEAATAGAOVLVPEPA--TMCRLGVPLRQVAPEVDPGPMANGVRRIATEA 132  
 QY 117 GIWLSLGFHFERGQDMONOKIYNCHVLL--NSKGSVVASIKTKHLCVETIRGQPMRS 174  
 |||||  
 DB 133 GITYAGMFTPTG-----DGRVTNLLIAGPCTPNQPDANHYKHLHYDA----FGFTES 182  
 QY 175 NYTRPGGTLEPPVKTPEG-KVGLAICYDMRPPELSKLTAOAGAEILUTYPSAFSGVTGPA- 232  
 |||||  
 DB 183 RTVARG--REPVVVVVVDVGRGLTYCYDIRPALYTELARAKQILAVCASMGs--GPGK 238  
 QY 233 --HMEVLLARAIESQCYVIAAQC-----GRHNETRASYSCHSNVVDPMGTIVVANC 281  
 |||||  
 DB 239 LEQWTLARARALDSMSVVAAGQADPGDARTGVGASSAAPTGVGSLVSGEVVSA 298  
 QY 282 SEGELCLARIDLHFLQMRQHLVFGHR 310  
 |||||  
 DB 299 GTPQLVADIDVDVNAARIRIAVLRLNQ 327  
 |||||  
 RESULT 8  
 YPOQ\_PSEFL STANDARD; PRT; 285 AA.  
 ID YPOQ\_PSEFL  
 AC P55176;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHEICAL 31.2 KDA PROTEIN IN P00F 5 REGION (ORF2).  
 OS Pseudomonas fluorescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 NCBI\_TaxID=294;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CHAO;  
 RX MEDLINE=96064397; PubMed=8526497;  
 RA Schnider U., Keel C., Defago G., Haas D.;  
 RT "Tns-directed cloning of pqg genes from Pseudomonas fluorescens CHAO:  
 RT mutational inactivation of the genes results in overproduction of the  
 RT antibiotic pyoluteorin";  
 RL Appl. Environ. Microbiol. 61:3856-3864(1995).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X87299; GAA60729.1; -  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR001110; UPF0012.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS01227; UPF0012; 1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 285 AA; 31163 MW; 68B7C64F38CBDECB CRC64;

Query Match 15.4%; Score 268; DB 1; Length 285;  
 Best Local Similarity 28.8%; Pred. No. 5.5e-17;  
 Matches 90; Conservative 50; Mismatches 123; Indels 50; Gaps 11;

QY 12 LCTGTRLLRLRYLVLTQPRPRTSSSTSWELPLVAVCQVSTPNKQENFKTCAELVOEAR 71  
 |||||  
 DB 14 LSVSGVTKTRVALYLQCPPRPLDVG-----NQRHLQVAMEAT 51  
 QY 72 RLGACLAFLPEAFDFTARNPAETLLSEPLNGDLLGOYSOLARECGIWSLGFHFERGQ 130  
 |||||  
 DB 52 --DADLLVPEMFSLSYNTGLEAVGALAEADQPSAQRITAAIAQAAGTAL-LGYIERSV 108  
 QY 131 DWEONOKIYNCHVLLNSKGSVVASIKTKHLCVETIRGQPMRSNTYKRGTELEPPVKT 190  
 |||||  
 DB 109 DQG---YINAVQLIDAQORLCNVRKTHLF-----GDLDSHMS-AGEDDFPLVEID 156  
 QY 191 AGKVLATIDYDMRPPELSKLTAOAGAEILUTYPSA---FGSYGPAHMEVLLARAIESQ 246  
 |||||  
 DB 157 GKKLGLFLCYDIEFENARRLALAGAEILLVPTANNIPYDEVA---DVTIRARAFENQ 211  
 QY 247 CVVIAAOCGRHNETRASV-GHSNVVDPMGTIVVAVRCSEGGCLARIDLHFLQMRQHL 305  
 |||||  
 DB 212 CYVAVANYCG--HEEDIRYCGOSSIAAPDSRIALAGLDEALITGLDROLMEESALNR 269  
 QY 306 VFOHRRPDLYGSL 318  
 |||||  
 DB 270 YLSDRRRELXDDL 282  
 |||||  
 RESULT 9  
 YAG5\_STAU STANDARD; PRT; 261 AA.  
 ID YAG5\_STAU  
 AC P55177;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHEICAL 29.8 KDA PROTEIN IN AGR OPERON (ORF 5).  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Staphylococcus.  
 NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ISOLATE GAL;  
 RX MEDLINE=96004766; PubMed=7565609;

RA Novick R.P., Projan S.J., Kornblum J., Ross H.F., Ji G.,  
 RA Kreiswirth B., Vandenesch F., Moghazeh S.:  
 RT "The agr P2 operon: an autocatalytic sensory transduction system in  
 RT *Staphylococcus aureus*.";  
 RL Mol. Gen. Genet. 248:446-458(1995).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X52543; CAA36779.1; -  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR001110; UPF0012.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS01227; UPF0012; 1.  
 KW Hypothetical protein.  
 SO SEQUENCE 261 AA; 29826 MW; 2303D95A8C18F838 CRC64;

Query Match 14.4%; Score 249.5; DB 1; Length 261;  
 Best Local Similarity 27.1%; Pred. No. 2.3e-15;  
 Matches 79; Conservative 51; Mismatches 109; Indels 53; Gaps 9;  
 QY 40 WEPLV-----AVCQYSTRPKQENKTCALVQEARLGCALFPE-----AFD 85  
 DB 6 YQPLVIGDSSKNETQTQTFWEKKNMAEVDVVLPEMNNNYDLEHINEKADNNGSFS 65  
 QY 86 FIARNPATELLSEPLNGDGL-GQYSOLARECGIWLISGFERHGDQWONKITYCHVL 144  
 DB 66 FIKH-----LAKKYVDIYAGSVSN-----RNNQIFNTAFS 97  
 QY 145 LNSKGSVAVSYRKTHTLCLDVEIPGQPMRESNYTRPGGLEPVPKTPACK-VGLAICYDMR 203  
 DB 98 VNSGQGLINEYDKVHLVPM-----LREHFLTAGEVVAEPDLSDTYVTLQICYDLR 150  
 QY 204 FPPELSKLAAQAGAILTPPSAFSGSVTPGAHWEVLLRRAITSCQCVYIAAOCGRHETRA 263  
 DB 151 FPELRLYPARSGAKIAYVQWQ-PMRSLQHWHSLLKARATENNMFVIGTNGFGDGTNEY 209  
 QY 264 SYGSHVVDPMGVAVARCSEGGCLARIDLHFLQOMKROHLPVQHRPRDY 315  
 DB 210 A-GHSIYINRGDLVGLNESADILTYVDNLINVEQORENIPVKSIXLDLY 260  
 RESULT 10  
 YAG5\_STPALU STANDARD; PRT; 234 AA.  
 AC P55178;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOHETICAL PROTEIN IN AGR OPERON (ORF 5) (FRAGMENT).  
 OS *Staphylococcus lugdunensis*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC *Bacillus/Staphylococcus* group; *Staphylococcus*.  
 CC NCBI\_TaxID=28035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9336158; Pubmed=8359673;  
 RA Vandenesch F., Projan S.J., Kreiswirth B., Etienne J., Novick R.P.:  
 RT "Agr-related sequences in *Staphylococcus lugdunensis*.";  
 RL FEMS Microbiol. Lett. 111:115-122(1993).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L13334; AAA71975.1; -  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR001110; UPF0012.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS01227; UPF0012; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SO SEQUENCE 234 AA; 26495 MW; C1CBADCA5E1389A7 CRC64;

Query Match 13.7%; Score 237.5; DB 1; Length 234;  
 Best Local Similarity 28.7%; Pred. No. 2.5e-14;  
 Matches 74; Conservative 37; Mismatches 118; Indels 29; Gaps 6;  
 QY 59 NFKTCALVQEARLGCALFPEAFDFIARNPATELLSEPLNGDLIGQYSOLARECGI 118  
 DB 4 NKQTDVVLPEMNNNGYALFQLEEKADFLERSDFI-----KNLALQYQV 49  
 QY 119 WLSIGFERHGDQWONKITYCHVLLNSKGSVAVSYRKTHTLCLDVEIPGQPMRESNYTK 178  
 DB 50 DIIAGSVSNKHH-----HIFNTAFIDTKGVINQYDKMLVPM-----LDEPAFLT 97  
 QY 179 PGCTLEPVPKTPAG-KVGLAICYDMRPELSKLAAQAGAILTPPSAFSGVTGAHWEVL 237  
 DB 98 AGKAVPEFKLSNGVKYQMTCYDLRPELRLYPARSGATIAFYAQMPSAR-LNHQVYL 156  
 QY 238 LRAARISQCVYIAAOCGRHETRAFGSHMVDPMGVAVARCSEGGCLARIDLHFL 297  
 DB 157 LKARAIENMVMYICNGCGYDKGTQYA-GHSVAIINPGEIITQELSTKEKELTVIDDAV 215  
 QY 298 QOMROHLPVQHRPRDY 315  
 DB 216 EQORRAIPVFDLSLVPHLY 233  
 RESULT 11  
 NRLL\_ARATH STANDARD; PRT; 346 AA.  
 ID NRLL\_ARATH  
 AC P32961; Q42543; O04908;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NITRILASE 1 (EC 3.5.5.1).  
 GN NIT1 OR AT3G44310 OR T10D17\_100.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 CC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RP STRAIN=CV. LANDSBERG ERECTA; TISSUE=leaf;  
 RX MEDLINE=92209532; Pubmed=155601;  
 RA Bartling D., Seedorf M., Mithoefer A., Weller E.W.:  
 RT "Cloning and expression of an Arabidopsis nitrilase which can convert  
 RT indole-3-acetonitrile to the plant hormone, indole-3-acetic acid.";  
 RL Eur. J. Biochem. 205:417-424(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Zhou L., Bartel B., Thorburn R.W.:  
 RT "Nucleotide sequence of the Arabidopsis thaliana nitrilase 1 gene.";  
 RL (in) Plant Gene Register PGR95-130.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98145459; Pubmed=9484465;  
 RA Hillebrand H., Bartling D., Weller E.W.:  
 RT "Structural analysis of the nit2/nit1/nit3 gene cluster encoding  
 RT nitrilases, enzymes catalyzing the terminal activation step in indole-



RT acetic acid biosynthesis in *Arabidopsis thaliana*;  
 RL Plant Mol. Biol. 36:89-99(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016720; PubMed=11130713;  
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unsel M.,  
 RA Fattmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Deleney M., Boutry M., Griwell L.A., Maché R., Puigdomenech P.,  
 RA De Simone V., Choise N., Artiguenave F., Robert C., Brotlier P.,  
 RA Wincker P., Catolico L., Weissbach J., Saurin W., Quettier F.,  
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Mumbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,  
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T., Nordstiek G.,  
 RA Reichelt J., Scharfe M., Schoen O., Barques M., Terol J., Climent J.,  
 RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,  
 RA Cooke R., Laurie M., Berger-Llauro C., Purnelle B., Masny D.,  
 RA de Haan M., Maare A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
 RA Monfort A., Argilhon A., Flores M., Liguori R., Vitale D.,  
 RA Mannhaupt G., Haese D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Walts A., Uteback T., Fujii C.Y., Shea T.P.,  
 RA Cressy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Milischer J., Sellers P., Gill J.E., Feldlyum T.V.,  
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asami E.,  
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsunoto M., Matsuno A., Muraki A.,  
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.,  
 RT "Sequence and analysis of chromosome 3 of the plant *Arabidopsis*  
 RT *thaliana*.";  
 RL Nature 408:820-822(2000).  
 [5]  
 RP CHARACTERIZATION  
 RC STRAIN=CV. LANDSBERG. ERCTA;  
 RX MEDLINE=94286570; PubMed=8016109;  
 RA Bartling D., Seedorf M., Schmidt R.C., Weiler E.W.;  
 RT "Molecular characterization of two cloned nitrilases from *Arabidopsis*  
 RT *thaliana*: key enzymes in biosynthesis of the plant hormone indole-3-  
 RT acetic acid";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).  
 CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE  
 CC INDOLE-3-ACETIC ACID.  
 CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, BUT AT A  
 CC VERY LOW LEVEL DURING THE FRUITING STAGE.  
 CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X63445; CAA45041.1; -;  
 DR EMBL: U38845; AAB05221.1; -;  
 DR EMBL: Y07648; CAA68935.2; -;  
 DR EMBL: AL353865; CAB88999.1; -;  
 DR PIR: S22398; S22398.  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR000132; Nitril\_cyn\_hydrolase.  
 DR Pfam: PF00795; CN\_hydrolase\_1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
 KM Hydrolase: Multigene family.  
 FT ACT\_SITE 186 186 BY SIMILARITY.  
 FT CONFLICT 312 312 Y -> H (IN REF. 2).  
 FT SEQUENCE 346 AA; 38178 MW; 8DAF887CAD1E3C1F CRC64;

Query Match 11.9%; Score 206; DB 1; Length 346;  
 Best Local Similarity 25.7%; Pred. No. 2.8e-11;  
 Matches 86; Conservative 47; Mismatches 117; Indels 84; Gaps 14;  
 QY 27 TOPRPTMSSTWELPVAVCOVTSPPNKEPFCALVEGARALAPLPEF-- 84  
 DB 23 TVRTVTVOSTVYN-----DTPATIDAEKY-----IVRASKAGELVLPESGFIG 69  
 QY 85 -----DFIARNPAETLLSEPLNDLGOYSOLARECIGWISLIG 124  
 DB 70 GYPRGFRGLAVGNHEGDEFRKYHASAI-----HYPGGEVARLADVARKNNVYLWGA 125  
 QY 125 FHERGDMEQNKIVNCHVLLNS-KGSVASYKTKLGDVE--IPGGPMRESNYTKPGG 181  
 DB 126 IEKEG-----YLYL-CTVLFSPGCGFLCKHKKLMPSTLERICWCG-----DG 168  
 QY 182 TLEPPYKTPAGKVGALICVYMRPELSLKAQGAELITPSPAGSVTGPANHEVLLRAR 241  
 DB 169 STIPVYDTPIGKLGAAICWENRMPLYRTALYARGIELYCAPADGS---KEQSSMLHT 224  
 QY 242 AHSOCGVVIAAO-CGR-----HHETRASYGSHSVYDPMGTVVARCS 282  
 DB 225 AITGGCGFVLSACQFCQKRPDPDLFTWYDKEHDSIVSGGYSIISPLGVLAPN 284  
 QY 283 -EGPGCLARIDLHFLQGRHQL-PVFGHRRPDL 314  
 DB 285 FESEGLVTADIDGLDIARAKLYFDSVGYSPDPV 318  
 RESULT 12  
 NR13\_ARATH STANDARD; PRT; 346 AA.  
 ID NR13\_ARATH  
 AC P46010; O04909;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NITRILASE 3 (EC 3.5.5.1).  
 GN NIT3 OR AT3G44320 OR T10D17\_110.  
 OS *Arabidopsis thaliana* (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; *Arabidopsis*.  
 OC NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=94294436; PubMed=8022831;  
 RA Bartel B., Fink G.R.;  
 RT "Differential regulation of an auxin-producing nitrilase gene family  
 RT in *Arabidopsis thaliana*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=98145459; PubMed=9484465;  
 RA Hillebrand H., Bartling D., Weiler E.W.;  
 RT "Structural analysis of the nitr/nitril gene cluster encoding  
 RT nitrilases, enzymes catalyzing the terminal activation step in indole-  
 RT acetic acid biosynthesis in *Arabidopsis thaliana*.";  
 RL Plant Mol. Biol. 36:89-99(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016720; PubMed=11130713;  
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unsel M.,  
 RA Fattmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Deleney M., Boutry M., Griwell L.A., Maché R., Puigdomenech P.,  
 RA De Simone V., Choise N., Artiguenave F., Robert C., Brotlier P.,  
 RA Wincker P., Catolico L., Weissbach J., Saurin W., Quettier F.,  
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Mumbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,

RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,  
 RA Reichelt J., Scherfe M., Schoen O., Barques M., Terol J., Clement J.,  
 RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,  
 RA Cooke R., Lande M., Berger-Liauro C., Furnelle B., Masuy D.,  
 RA de Haan M., Maestre A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
 RA Monfort A., Argitlou A., Flores M., Lignori R., Vitale D.,  
 RA Mannhaupt G., Haese D., Schoof H., Rudd S., Zaccaria P., Meves H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Maltis A., Uterback T., Fujil C.Y., Shea T.P.,  
 RA Cressy T.H., Haas B., Maltis R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Miltscher J., Sellers P., Gill J.E., Feldlyum T.V.,  
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 RA Kiyokawa S., Kohara M., Matsuno M., Matsuno A., Muraki A.,  
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Matanabe A., Yamada M., Yasuda M., Tabata S.,  
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:820-822(2000).  
 CC -I- FUNCTION: CAN CONVERT INDOL-3-ACETONITRILE TO THE PLANT HORMONE  
 CC INDOL-3-ACETIC ACID.  
 CC -I- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).  
 CC -I- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.  
 CC -I- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; U09959; AAA19627.1; -;  
 DR EMBL; Y07648; CA68936.2; -;  
 DR EMBL; AL353865; CAB89000.1; -;  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR000132; Nitril\_cyn\_hydrolase.  
 DR Pfam; PF00795; CN\_hydrolase; 1.  
 DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
 DR HydroLase; MultiGene family.  
 FT ACT\_SITE 186  
 FT SITE 186  
 FT SEQUENCE 346 AA; 38022 MW; 70CFF421547F2B5E CRC64;

Query Match 11.7%; Score 203; DB 1; Length 346;  
 Best Local Similarity 24.7%; Pred. No. 5.3e-11;  
 Matches 82; Conservative 50; Mismatches 116; Indels 84; Gaps 13;

QY 30 RPRMTSSSTSWELPVAACQVSTPNKQENFTCAELVQEARLACLAFLDEAF----- 84  
 DB 26 RVTIIVSSSTVYN-----DTPATLDKAEK-----IVEAASKAKVLLEFAEAGVGP 72  
 QY 85 -----DFIARNPAETLLSEPLNGDLGOYOLARCGIWLSTGPH 126  
 DB 73 RGFREGIAGVYNEGRDFE--RNYHASAI--KVPGEVELELAELAGNNVHLVGALE 127  
 QY 127 ERGDMEONQKTYNCVHLNKGVSVAASYRKTILCDVE--IPGQPMRESNYTKPGTLE 184  
 DB 128 KDG-----YTLCTALFFSPGQFLGKHKRYMPTSLERCINGOG-----DGSIT 171  
 QY 185 PPVKTPACKVGLATCYDMARFPELSIKLAQAGAEILTPSAFSSVYGRAPWEVLLARALE 244  
 DB 172 PYVDPIPKIGALIGEMENRMPYRTALYAKGIEICAPTADSL-----EWOASMIHIAE 227  
 QY 245 SOCYVIAAO--CGR-----HHEFRASVGSYMWVDPKGVAAVACS--EG 284  
 DB 228 GGCFFVLSHQCKRREFPEHDPYLFENDIVDTKEHPTVSGGGSVITISPLGKVLAPNYES 287

QY 285 PGLICARIDLHFLQONRQHLVPFOH--RRPDLY 315  
 DB 288 ESLVATADLDGIARAKLYFDVGVHYSKPDIF 319  
 RESULT 13  
 BUP\_RAT  
 ID BUP\_RAT STANDARD; PRT; 393 AA.  
 AC 003248;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE BUP-UREIDOPROPIONASE (EC 3.5.1.6) (BETA-ALANINE SYNTHASE)  
 OS (N-CARBAMOYL-BETA-ALANINE AMIDOHYDROLASE).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 202-212.  
 RC TISSUE=Liver;  
 RX MEDLINE=93194866; PubMed=8449931;  
 RA "Kvalnes-Krick K.L., Traut T.W.;  
 RT "Cloning, sequencing, and expression of a cDNA encoding beta-alanine  
 RT synthase from rat liver.";  
 RL J. Biol. Chem. 268:5686-5693(1993).  
 CC -I- CATALYTIC ACTIVITY: N-CARBAMOYL-BETA-ALANINE + H(2)O =  
 CC BETA-ALANINE + CO(2) + NH(3).  
 CC -I- COFACTOR: BINDS TWO ZINC IONS.  
 CC -I- ENZYME REGULATION: ALLOSTERIC ENZYME WITH POSITIVE COOPERATIVITY  
 CC TOWARD THE SUBSTRATE N-CARBAMOYL-BETA-ALANINE.  
 CC -I- PATHWAY: THIRD AND FINAL STEP IN THE CATABOLISM OF THE  
 CC PYRIMIDINE BASES, URACIL OR THYMINE, TO PRODUCE BETA-ALANINE OR  
 CC 2-METHYL-BETA-ALANINE.  
 CC -I- SUBUNIT: IN THE ABSENCE OF LIGANDS, THE ENZYME EXISTS AS A STABLE  
 CC HOMODIMER, ALTHOUGH THIS IS NOT THE MOST ACTIVE FORM OF THE  
 CC ENZYME. THIS NATIVE HEXAMER DISSOCIATES TO AN INACTIVE TRIMER IN  
 CC RESPONSE TO THE PRODUCT, BETA-ALANINE, OR ASSOCIATES TO THE MORE  
 CC ACTIVE HOMODIMER IN RESPONSE TO THE SUBSTRATE.  
 CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; M97662; AAA40804.1; -;  
 DR PIR; S27881; S27881.  
 DR PIR; A46624; A46624.  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR Pfam; PF00795; CN\_hydrolase; 1.  
 DR HydroLase; Zinc.  
 FT MOD\_RES 71  
 FT MOD\_RES 71  
 FT METAL 97  
 FT METAL 97  
 FT METAL 101  
 FT METAL 101  
 FT METAL 158  
 FT METAL 158  
 FT METAL 280  
 FT METAL 280  
 FT METAL 293  
 FT METAL 293  
 FT METAL 297  
 FT METAL 297  
 FT SEQUENCE 393 AA; 44042 MW; 723E5AF5B01E3AAD CRC64;

Query Match 11.1%; Score 193; DB 1; Length 393;  
 Best Local Similarity 24.7%; Pred. No. 5e-10;  
 Matches 71; Conservative 40; Mismatches 147; Indels 30; Gaps 6;

QY 51 TSTPNKOE--NFKCAELVQEARLGLACTIAPLEA-----PFIIARNPAETLLSEPLNG 103  
 DB 85 TSAPVABOVSALHKRIETAEVAAACGVNITCFQEAAMMPFACTREKLPWTEFAESAD 144

OY 104 DLGGYSLARECGIWLSTGFGHERGQDWEQNOIKYCNHVLNLSKSGSVVAYSRKTHLCDV 163  
 DB 145 GILTRRCQKLAKKHNWVSPILERRD--HGVLMNTAVINSGLVMKTKTKNH---- 198  
 OY 164 ETPGQPMRESNRYKGGTLEPPVKTTPACKVGLAICYDMRPELSKLQAQAEILTYPS 223  
 DB 199 -IPRVDPMESNYVMGNLGHFVQFOGRIANVICYGRHHPHLMVSYNGAEIIFNPS 257  
 OY 224 AGSGYVGRHWEVLARARIESOCYIAAAGCR-----HHETRASVGH 268  
 DB 258 ATGELSESMWPLEARNAAIAHCFCAINRGQEHYPNEFTSGDKKAAHDLGYEYSS 317  
 OY 269 MYVDPGTVVARSCEG-PEGLTARIDLHFLQWRQHLPFVQHRPPDY 315  
 DB 318 YVAAPDGRTPGLSRNQDGLVLTALNLCQOINDPWTFRMTGRLEMY 365  
 RESULT 14  
 NR12 ARATH STANDARD: PRT: 339 AA.  
 AC P32962: Q96505; (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NITRILASE 2 (EC 3.5.5.1).  
 GN NIT2 OR AT3G44300 OR T10D17\_90.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. LANDSBERG. ERECTA; TISSUE=Leaf;  
 RX MEDLINE=94286570; PubMed=8016109;  
 RA Bartling D., Seedorf M., Schmidt R.C., Weiler E.W.;  
 RT "Molecular characterization of two cloned nitrilases from Arabidopsis  
 thaliana: key enzymes in biosynthesis of the plant hormone indole-3-  
 acetic acid.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=94294436; PubMed=8022831;  
 RA Bartel B., Fink G.R.;  
 RT "Differential regulation of an auxin-producing nitrilase gene family  
 in Arabidopsis thaliana.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Zhou L., Bartel B., Thornburg R.W.;  
 RT "Nucleotide sequence of a pathogen induced nitrilase gene from  
 Arabidopsis thaliana: Nlt2.";  
 RL (In) Plant Gene Register PGR96-006.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016720; PubMed=11130713;  
 RA Farmanbat M., Lemcke K., Rieger M., Ansoorge W., Unseld M.,  
 RA Salmannat M., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,  
 RA Deleseny M., Boutry M., Griwall L.A., Mache R., Puigdomenech P.,  
 RA Simone V., Choise N., Artiguenave F., Robert C., Brotlier P.,  
 RA Winkler P., Catolico L., Weissendach J., Saurin W., Queller F.,  
 RA Schuber M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Wurnbach E., Drzonek H., Erle H., Jordan N., Bangert S.,  
 RA Wieselmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T.H., Nordsiek G.,  
 RA Reichelt J., Scharte M., Schoen O., Barques M., Terol J., Clement J.,  
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,  
 RA Cooke R., Laude M., Berger-Liandro C., Purnelle B., Masuy D.,  
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,

RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,  
 RA Manhaupt G., Haase D., Schoof H., Rued S., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P.,  
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Miltcher J., Sellers P., Gill J.E., Feldblyum T.V.,  
 RA Paus D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato T., Asanizu E.,  
 RA Sasamoto S., Kimura T., Ideasa K., Kawashima K., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakayama S., Nakazaki N., Shimizu S., Takeuchi C., Wada T.,  
 RA Matanabe A., Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 408:820-822(2000).  
 CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE  
 CC INDOLE-3-ACETIC ACID.  
 CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).  
 CC -1- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.  
 CC -1- DEVELOPMENTAL STAGE: BARELY DETECTABLE IN YOUNG ROSETTES, BUT IS  
 CC STRONGLY EXPRESSED DURING BOLTING, FLOWERING, AND ESPECIALLY FRUIT  
 CC DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: X68305; CAA48377.1; -;  
 CC EMBL: U09958; AAB60275.1; -;  
 CC EMBL: U38845; AAB05220.1; -;  
 CC EMBL: AL353865; CAB8998.1; -;  
 CC PIR: S31969; S31969.  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR000132; Nitril\_cyn\_hydrolase.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
 KW Hydrolase; Multigene family.  
 FT ACT\_SITE 179 179 BY SIMILARITY.  
 FT CONFLICT 37 37 E -> G (IN REF. 3).  
 FT CONFLICT 48 48 S -> T (IN REF. 3).  
 FT SEQUENCE 339 AA; 37153 MW; 06CDE76D2FDC24A7 CRC64;  
 Query Match 10.88; Score 187.5; DB 1; Length 339;  
 Best Local Similarly 24.28; Pred. No. 1.3e-09;  
 Matches 79; Conservative 51; Mismatches 124; Indels 73; Gaps 12;  
 OY 34 MSSSTWELPLVAVCOY-TSPNKGKQENKTCALVEARAGACLAFLPEAF----- 84  
 DB 12 VASSTYIRATIVVASTYVNDTPATLEK--ANKTIVRAASGSELVYPPEAFICGRGF 68  
 OY 85 -----DFTARNPAETLLSEPLNGLGOYSLARECGIWLSTGFGHERGQ 130  
 DB 69 RFGLVGVHNEDEGRDEPRKYHASAI-----KVPGEVEKLAEKNNVYLWGAIEKDG- 123  
 OY 131 DWEDNOKIYCNHVLNLSKSGSVVASYRKTTHLCDVE--IPGQPMRESNRYKGGTLEPPV 188  
 DB 124 -----VTLCTALFEFSQGGFLGHRKRLMPTSLRRCITWGG-----DGSTIPYD 168  
 OY 189 TPAGKVLATCYDMRPELSKLQAQAEILTYPSAGSVYVPAHWEVLARARIESOCY 248  
 DB 169 TRPICKLAALCWMEMRLYRATYAKGIELYCAPTADG-----KEMSSMHNHIEGCGF 224  
 OY 249 VVAAAQ-C-----GRNHETRASVGHMYVDPGTVVARS-EGFGLC 288  
 DB 225 VLSACQCLRKDPFDHDPDYLTDMYDDEKPSIVSOGSGVITSLGVLGAPNPESEGLI 284

OY 289 LARIDLHFLQOMRQHL-PVFGHRRPDL 314  
 DB 285 TADLDLDVAPRAKLYEDSVGHYSRPDV 311

## RESULT 15

ALAM\_PSEAE STANDARD: PRT; 346 AA.

AC P11436;

DT 01-OCT-1989 (Rel. 12, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

GN AMIE OR PA3366.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

CC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE.

RC STRAIN=PA3142;

RX MEDLINE=87219101; PubMed=3108029;

RA Ambler R.P., Auffret A.D., Clarke P.H.;

RT "The amino acid sequence of the aliphatic amidase from Pseudomonas aeruginosa.";

RL FEBS Lett. 215:285-290(1987).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=87219102; PubMed=3108030;

RA Brummel W.J., Charles I.G., Matfield M., Liu C.-P., Drew R.E.,

RA Clarke P.H.;

RT "The nucleotide sequence of the amle gene of Pseudomonas aeruginosa.";

RL FEBS Lett. 215:291-294(1987).

RN [3]

RP SEQUENCE FROM N.A.

RX STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";

RL Nature 406:959-964(2000).

CC -!- FUNCTION: ALLOWS THE ORGANISM TO GROW ON SHORT CHAIN-LENGTH ALIPHATIC AMIDES.

CC -!- FUNCTION: ENABLES THE ORGANISM TO USE ACETAMIDE AS BOTH CARBON AND A NITROGEN SOURCE.

CC -!- CATALYTIC ACTIVITY: A MONOCARBOXYLIC ACID AMIDE + H(2)O = A MONOCARBOXYLATE + NH(3).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M27612; AAA25697.1; -.

DR EMBL; AE004759; AAG06754.1; -.

DR PIR; A26741; A26741.

DR InterPro; IPR003010; CN\_hydrolase.

DR Pfam; PF00795; CN\_hydrolase; 1.

KW Hydrolase; Complete proteome.

FT CONFLICT 40 40 E -> D (IN REF. 1 AND 2).

FT CONFLICT 169 169 G -> P (IN REF. 2).

FT CONFLICT 317 317 E -> D (IN REF. 1 AND 2).

SO SEQUENCE 346 AA; 38495 MW; E19CEB474EB92B93 CRC64;

Query Match 10.7%; Score 186.5; DB 1; Length 346;  
 Best Local Similarity 25.4%; Pred. No. 1.7e-09;  
 Matches 73; Conservative 44; Mismatches 113; Indels 57; Gaps 14;

OY 56 KOENFTCAELVQEARL-----GACLAFLPE-AEDFARNAEITLLSEPTL 102  
 DB 21 KMERLHTAAEVLDMARKIAEMIVGMKQGLPMDLVFPEYSLQGIWYDPAEMMETAVAI 80  
 OY 103 GDLLGOYSLARCGIWLISLGFHERGQDMQONOK--IYNCHVILSKGSVVAAYRK-TH 159  
 DB 81 GEETELFSRACKRANW---GVPSLTGERHEHPRAPNTLVLDINNGEIVQYKRIIP 137  
 OY 160 LCDVE--IPGQPMRESNYTKPGTLEPPVKTPAG-KVGLAICYDMRFPFLSLKAQAQA 216  
 DB 138 WCPLEGWYPG-----GOTVYSEG-----PKGKISLILICDDGNYPDEIWRDCAMKA 183  
 OY 217 EILT-----YSARSVYGRPAHWEVLLARALESCYVIAAAGCRHHETASYSHSW 270  
 DB 184 ELIVRCGYWYPAKQOV-----MMARAMANNQYV-AVANAAAGPDGVVSYFGSHA 235  
 OY 271 VDPWGTIVARCEGP-GLCLARIDLHFLQ-----OMROHLPVFOHR 310  
 DB 236 IGFDTGLGECGEEEMGIYQAQLSLQIRDAANDQSQNHLEKILHR 282

Search completed: April 26, 2002, 17:02:50  
 Job time: 453 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 17:02:22 ; Search time 57.61 Seconds

(without alignments)  
820.100 Million cell updates/sec

Title: US-09-357-675C-22

Perfect score: 1736

Sequence: 1 MGFTRPPHQLCTGYRL.....LPVQHRPRDLXGSLGHPIS 323

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp\_invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1736	100.0	323	11	088526
2	1731	99.7	323	11	0981M4
3	1554	89.5	290	11	098B22
4	1487	85.7	327	4	076091
5	921	53.1	288	13	0918G0
6	706	40.7	440	5	076463
7	699.5	40.3	460	5	076464
8	632.5	36.4	316	10	091E50
9	588.5	33.9	276	3	094660
10	538.5	31.0	272	1	027839
11	491	28.3	270	2	083040
12	483	27.8	276	4	09NORA
13	482	27.8	282	2	09HVU6
14	470.5	27.1	276	11	09JHW2
15	431.5	24.9	275	2	09KUD4
16	427.5	24.6	283	2	09A9Y3
17	417	24.0	318	10	09LYH1
18	410	23.6	270	2	09KOW5
19	404	23.3	270	2	09J713

20	393.5	22.7	189	11	09CTG9	09ctg9 mus musculu
21	389.5	22.4	283	5	09VHE4	09vhe4 drosophila
22	366	21.1	231	11	09D0S1	09d0s1 mus musculu
23	342.5	19.7	113	6	029367	029367 sus scrofa
24	341.5	19.7	264	2	09XA70	09xa70 streptomyce
25	340	19.6	257	1	030121	030121 archaeoglob
26	334	18.2	297	2	09RFF7	09rff7 deinococcus
27	332	18.5	270	1	09H1W8	09h1w8 thermoplas
28	290.5	16.7	300	10	09XG19	09xg19 lycopersico
29	285.5	16.4	292	2	0916J8	0916j8 pseudomonas
30	276.5	15.9	259	2	031664	031664 bacillus su
31	276.5	15.9	271	2	09KE11	09ke11 bacillus ha
32	275	15.8	287	1	09Y9L1	09y9l1 aeropyrum p
33	271	15.6	270	2	091241	091241 pseudomonas
34	262.5	15.1	220	10	09Z0H4	09z0h4 arbidopsi
35	262.5	15.1	262	2	09L104	09l104 streptomyce
36	262.5	15.1	280	2	069939	069939 streptomyce
37	262.5	15.1	295	2	09PA02	09pa02 xylella fas
38	261.5	15.1	292	2	09AB15	09ab15 caulobacter
39	257.5	14.8	282	2	09Z5U5	09z5u5 zymomonas m
40	255.5	14.7	280	2	069808	069808 streptomyce
41	255	14.7	298	12	089413	089413 paramacium
42	251.5	14.5	292	2	025079	025079 helicobacte
43	250.5	14.4	294	2	09ZMC7	09zmc7 helicobacte
44	249.5	14.4	261	2	099SL3	099sl3 staphylococ
45	248	14.3	292	2	025452	025452 helicobacte

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	323 AA.
088526	088526			
AC	088526			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	NITRILASE HOMOLOG 1.			
GN	NITL.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98337986; PubMed=9671749;			
RA	Pekarsky Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,			
RA	Tillib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,			
RA	Buchberg A.M., Maro A., Brenner C., Croce C.M.,			
RT	"Nitrilase and Fhl homologs are encoded as fusion proteins in			
RT	Drosophila melanogaster and Caenorhabditis elegans."			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).			
DR	EMBL: AF069985; AAC40184.1; -			
DR	MGI: MGI:1350916; Nitl.			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	InterPro: IPR001110; UPF0012.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPF0012; 1.			
SO	SOURCE 323 AA; 35717 MW; 10151CER15IDFZC7 CRC64;			

Query Match 100.0%; Score 1736; DB 11; Length 323;  
Best Local Similarity 100.0%; Pred. No. 5.3e-151;  
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGFTTRPPHQLCTGYRLRIPVCTQPPRMTSSSTWELPYAVCOVTPPNKQENF	60
DB	1	MGFTTRPPHQLCTGYRLRIPVCTQPPRMTSSSTWELPYAVCOVTPPNKQENF	60
QY	61	KTCAELVQEARLQACIAFLPEAFDFIARNPAETILLSEPLNGDLGQYSQLARECGIWL	120
DB	61	KTCAELVQEARLQACIAFLPEAFDFIARNPAETILLSEPLNGDLGQYSQLARECGIWL	120

OY 121 SLGGFHERGQDMEONOKITYNCHVLLNSKGSVYASRYKTHLCVETIPGSGPMRESNTTKRG 180  
DB 121 SLGGFHERGQDMEONOKITYNCHVLLNSKGSVYASRYKTHLCVETIPGSGPMRESNTTKRG 180  
OY 181 GTLEPPVKTTPACKVGLAICYDMRPFELSLKLAQAGAEILTPSARGSVTPGAHWEVLLRA 240  
DB 181 GTLEPPVKTTPACKVGLAICYDMRPFELSLKLAQAGAEILTPSARGSVTPGAHWEVLLRA 240  
OY 241 RAIESQCYVIAAOCGRHNETRASVGHSMVNDPVGCTVVARCSEGGELCLARIDLHFLQOM 300  
DB 241 RAIESQCYVIAAOCGRHNETRASVGHSMVNDPVGCTVVARCSEGGELCLARIDLHFLQOM 300  
OY 301 ROHLPVFOHRRPDLVGLSGLHPLS 323  
DB 301 ROHLPVFOHRRPDLVGLSGLHPLS 323

RESULT 2  
O9R1N4 PRELIMINARY: PRT: 323 AA.  
ID O9R1N4  
AC O9R1N4  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE NITRILASE 1.  
GN NIT1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98337986; PubMed=9671749;  
RA Pekarsky Y., Campigilio M., Siprashtvili Z., Druck T., Sedkov Y.,  
RA Tilib S., Draganescu A., Wernuth P., Rothman J.H., Huebner K.,  
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;  
RT "Nitrilase and flit homologs are encoded as fusion proteins in  
RT Drosophila melanogaster and Caenorhabditis elegans."  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
DR EMBL: AF069988; AAC40185.1; -.  
DR MGD: MGI:1350916; Ntl1.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS01227; UPF0012; 1.  
SQ SEQUENCE 323 AA; 35705 MW; F8CD7730713665EF CRC64;

Query Match 99.7%; Score 1731; DB 11; Length 323;

Best Local Similarity 99.7%; Pred. No. 1.5e-150;

Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MIGEITRPPHOLLCTGYRLRLIPVLCOTOPRPTMSSSTSWELPLVAVCOVSTPNKQENF 60  
DB 1 MIGEITRPPHOLLCTGYRLRLIPVLCOTOPRPTMSSSTSWELPLVAVCOVSTPNKQENF 60  
OY 61 KTCALVQEAARLGAACLAFLPRAFDIFIANRPAETLLSPLNGDLLGOYSOLAREGIML 120  
DB 61 KTCALVQEAARLGAACLAFLPRAFDIFIANRPAETLLSPLNGDLLGOYSOLAREGIML 120  
OY 121 SLGGFHERGQDMEONOKITYNCHVLLNSKGSVYASRYKTHLCVETIPGSGPMRESNTTKRG 180  
DB 121 SLGGFHERGQDMEONOKITYNCHVLLNSKGSVYASRYKTHLCVETIPGSGPMRESNTTKRG 180  
OY 181 GTLEPPVKTTPACKVGLAICYDMRPFELSLKLAQAGAEILTPSARGSVTPGAHWEVLLRA 240  
DB 181 GTLEPPVKTTPACKVGLAICYDMRPFELSLKLAQAGAEILTPSARGSVTPGAHWEVLLRA 240  
OY 241 RAIESQCYVIAAOCGRHNETRASVGHSMVNDPVGCTVVARCSEGGELCLARIDLHFLQOM 300  
DB 241 RAIESQCYVIAAOCGRHNETRASVGHSMVNDPVGCTVVARCSEGGELCLARIDLHFLQOM 300

OY 301 ROHLPVFOHRRPDLVGLSGLHPLS 323  
DB 301 ROHLPVFOHRRPDLVGLSGLHPLS 323

RESULT 3  
O9DBF2 PRELIMINARY: PRT: 290 AA.  
ID O9DBF2  
AC O9DBF2  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE NITRILASE 1.  
GN NIT1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=LIVER;  
MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochia H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stabul F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Blake J., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Sakai K., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai J.H., Kohlsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
DR EMBL: AK004988; BAB33723.1; -.  
DR MGD: MGI:1350916; Ntl1.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS01227; UPF0012; 1.  
SQ SEQUENCE 290 AA; 31886 MW; 34493DBFE2170C71 CRC64;

Query Match 89.5%; Score 1554; DB 11; Length 290;

Best Local Similarity 100.0%; Pred. No. 2.2e-114;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 MSSSTSMELPLVAVCOVSTPNKQENFKCALVQEAARLGAACLAFLPRAFDIFIANRPAE 93  
DB 34 MSSSTSMELPLVAVCOVSTPNKQENFKCALVQEAARLGAACLAFLPRAFDIFIANRPAE 93  
OY 94 TLLSLEPLNGDLLGOYSOLARECGIMLSLGGFHERGQDMEONOKITYNCHVLLNSKGSVYA 153  
DB 94 TLLSLEPLNGDLLGOYSOLARECGIMLSLGGFHERGQDMEONOKITYNCHVLLNSKGSVYA 153  
OY 154 STRKTHLCVETIPGSGPMRESNTTKRGITLPPVKTTPACKVGLAICYDMRPFELSLKLAQ 213  
DB 154 STRKTHLCVETIPGSGPMRESNTTKRGITLPPVKTTPACKVGLAICYDMRPFELSLKLAQ 213  
OY 214 ACAGELITPSARGSVTPGAHWEVLLRARIESQCYVIAAOCGRHNETRASVGHSMVNDP 273  
DB 214 ACAGELITPSARGSVTPGAHWEVLLRARIESQCYVIAAOCGRHNETRASVGHSMVNDP 273  
OY 274 WGTVVARCSEGGELCLARIDLHFLQOMROHLPVFOHRRPDLVGLSGLHPLS 323  
DB 274 WGTVVARCSEGGELCLARIDLHFLQOMROHLPVFOHRRPDLVGLSGLHPLS 323



```

Db          24.1  MGVYVARSSEGGICLARIIDLHFLQDMQHLPVFGHRRPDLYGSLGHPLS 290

RESULT      4
076091      PRELIMINARY:      PROT:      327 AA.

AC          076091
AC          076091
DT          01-NOV-1998 (TREMBLrel. 08, Created)
DT          01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT          01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE          NITRILASE HOMOLOG 1.
GN          NIT1.
OS          Homo sapiens (Human).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX          NCBI_TaxID=9606;
RN          11
RP          MEDLINE=9837986; PubMed=9671749;
RX          Bekarsky Y., Campigillo M., Siprashvili Z., Druck T., Sedkov Y.,
RA          Tilibid S., Draganescu A., Wernerth P., Rothman J.H., Huebner K.,
RA          Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT          "Nitrilase and Flit homologs are encoded as fusion proteins in
RT          Drosophila melanogaster and Caenorhabditis elegans.";
RL          Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
RL          EMBL; AF069984; AAC39901.1; -
DR          EMBL; AF069987; AAC39907.1; -
DR          InterPro: IPR001110; UPP0012.
DR          InterPro: IPR003010; CN_hydrolase.
DR          Pfam; PF00795; CN_hydrolase_1.
DR          PROSITE; PS01227; UPP0012; 1.
SQ          SEQUENCE 327 AA; 35896 MW; 90CFEB9D4BA627B1 CRC64;

```

Query Match	Similarity	85.7%	Score 1487	DB 4	Length 327
Best Local	Similarity	84.4%	Pred. No. 3,4e-128		
Matches 276	Conservative	27	Mismatches 20	Indels 4	Gaps 2
OY	1	MIGFTTRPPHQ--LILCGYRLLRIRIPVLTQORPRFMS--SSTWELPLVAACGVSTPNK	56		
Db	1	MIGFTTRPPHRLSLILCGRLRIPQLSVLCAOORPRPMAISSSCCEPLVAACGVISTPNK	60		
OY	57	QENFTKCAELVQEARLGACIAFLPEAEFDIARNPAETLLLEPLNGDLLGOYSOLAREC	116		
Db	61	QONFTKCAELVREARLGACIAFLPEAEFDIARDAETLHLEBPLGKLLLEYTOLARCC	120		
OY	117	GIMLSLGGFHNGGOWEDONOKTYNCHVLINSNGSVYASRKTHLCDVETPGGPRAESNY	176		
Db	121	GIMLSLGGFHNGGOWEOTOKTYNCHVLINSNGAVVATRKTHLCDVETPGGPRMESNS	180		
OY	177	TKPGSTLPPVYTPAGKVGIAICYUMRPPELSLTKLAQACAEILTYPSAFGSYTPAHMVEY	236		
Db	181	TMPGSTLSPSYTPAGKIGIACVDMRPPELSLALAQACAEILTYPSAFGSTITGPANMVEY	240		
OY	237	LLRAFAIESCCVYIAAOCGRHETRASGHSNMVDDPMGTVVAKSESGGLCLARDLNF	296		
Db	241	LLRAFAIEFGCVYVAAOCGRHHEKRAASYGHSNMVDDPMGTVVAKSESGGLCLARDLNF	300		
OY	297	LQMRQHLPVNOHRRPDLXSLGHPLS	323		
Db	301	LROLRRHLPVFOHRRPDLXGNLGHPLS	327		
RESULT	5				
Q918G0					
ID	Q918G0	PRELIMINARY	PRT	288	AA
AC	Q918G0				
DT	01-OCT-2000	(Tremblrel. 15, Created)			
DT	01-OCT-2000	(Tremblrel. 15, Last sequence update)			
DT	01-JUN-2001	(Tremblrel. 17, Last annotation update)			
DE	NIT PROTEIN 1.				
GN	NITL				
OS	Xenopus laevis (African clawed frog).				

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amniota; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
OC Xenophoridae; Xenopus.  
NCBI\_TaxID=8355;  
RN [1]  
RP  
RP SEQUENCE FROM N.A.  
RX PubMed=10959838;  
RA Pace H.C., Hodavadekar S.C., Draganescu A., Huang J., Bleganowski P.,  
RA Pekarsky Y., Croce C.M., Brenner C.;  
RT "Crystal structure of the worm Nitfhit Rosetta Stone protein reveals a  
RT Nit tetramer binding two Fhit dimers.",  
RL Curr. Biol. 10:907-917(2000).  
DR EMBL: AF284575; AAF87104.1; -;  
DR InterPro: IPR003010; CN\_hydrolase.  
DR InterPro: IPR001110; UPF0012.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS01227; UPF0012;  
SQ SEQUENCE 288 AA; 32061 MW; B3673A80609BF042 CRC64;

Query Match	53.1%;	Score 921;	DB 13;	Length 288;
Best Local Similarity	58.7%;	Pred. No. 2e-76;		
Matches 162;	Conservative 56;	Mismatches 58;	Indels 0;	Gaps 0;
OY	43	PLVAACQVSTSPNNKQENKTCALVEQFAARLGCACLAFLPEAFDFIARRNPATILLSEPLN	102	
		:     :     :     :     :     :     :     :     :     :     :		
DB	7	PLIACQMTSTSDKREKNATKCRSLRLREAAGRACVFLPEAFDYIGGSTLEETLSAESIH	66	
OY	103	GDLLQVQSOLARECGIWLSLGGFHEHSGDWEQNKIYNCHVLLNKGSVVASYRTHLCD	162	
		: :     :     :     :     :     :     :     :     :     :		
DB	67	GDITQRYQLAECSGLMTSLGFGHEKGPMDTDQRTISNHYVDNTHIVSYRRALHFD	126	
OY	163	VEIPQGGPRESENTRKPGCTLEPPKTTAGKYGALICTDMRPELSKLQAQGAIIITYP	222	
		: :     :     :     :     :     :     :     :     :     :		
DB	127	VDLQNGVSLRSESSPTLPGAEILRPTTSPAGKGLGVLCYDLRPEFSLLAQGAELIITYP	186	
OY	223	SAFGSVTGPANHEVILLRRAIETSCCYVLAACGCHHETRATSGSMVDPGTVVARS	282	
		:     :     :     :     :     :     :     :     :     :		
DB	187	SAFTLTTLGIANHEVILLRRAITGCTCYVVAAMAQDHNHNEKRISYGHAMVYDPMGLVIGQC	246	
OY	283	EGPGLCLARIDLHPLQOMRQHLPVFQHRRPDLYGLS	318	
		: :     :     :     :     :     :     :     :     :     :		
DB	247	EGTGICVAEIDIPYMERVARDMPVRHRRHTDLYGKI	282	
RESULT	6			
O76463		PRELIMINARY;	PRT;	440 AA.
ID	O76463			
AC	O76463;			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	NITRILASE AND FRAGILE HISTIDINE TRIAD FUSION PROTEIN NITPHIT.			
GN	NITPHIT OR Y56A3A.13.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabditidae;			
OC	Rhabditidae; Pelodertinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98337986; PubMed=9671749;			
RA	PEKARSKY Y., Campigilio M., Siprashvili Z., Druck T., Sedkov Y.,			
RA	Tilibid S., Draganescu A., Wernuth P., Rothman J.H., Huebner K.,			
RA	Buchberg A.M., Mazo A., Brenner C., Croce C.M.,			
RT	Nitrilase and Flii homologs are encoded as fusion proteins in			
RT	Drosophila melanogaster and Caenorhabditis elegans."			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Mathews L.,			
RL	Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			

[illegible]

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Doudin K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dudkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Garfield A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Hostin N.L., Harrey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostali D., Houston K.A., Howland T.U., Wei M.-H., Ibegwan C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.B., Kelchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mosherfi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.T.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Plitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,  
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtkas R., Teector C., Turner R., Ventler E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98337986; PubMed=9671749;  
RA Pekarsky Y., Campiglio M., Siprasvill Z., Druck T., Sedkov Y.,  
RA Tilib S., Draganescu A., Wermuth P., Rothman J.H., Huebner K.,  
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;  
RT "Nucleotide and flit homologs are encoded as fusion proteins in  
RT *Drosophila melanogaster* and *Cenochabditis elegans*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
DR EMBL: AE003467; AAF7347.1; -;  
DR EMBL: AF069989; AAC39137.1; -;  
DR HSSP: P49789; 2FIT.  
DR FlyBase: FBgn0024945; Nitfhl.  
DR InterPro: IPR00310; CN\_hydrolase.  
DR InterPro: IPR00310; HT.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR Pfam: PF01230; HIT; 1.  
DR PROSITE: PS00892; HIT; 1.  
SQ SEQUENCE 460 AA; 52231 MW; 81121A00BC337706 CRC64;

```

Query Match          40.3%; Score 639.5; DB 5; Length 460;
Best Local Similarity 50.2%; Pred. No. 7.1e-56;
Matches 140; Conservative 44; Mismatches 88; Indels 7; Gaps 3.
```

Oy	45	VAVCVQVTPSPNQENFKCAELVOEARRAGACLAFLPEAFDFIARNPATLLSLEPTD	104
Db	35	IAVGMSRSTSDAAINISOVIELVDRRKSSONACMLPLPECDDVGESRTOFTLSBGLGE	94
Oy	105	LLGGYSOLARECGIMSLSGFERHGQDWEOBKOTICYNHVLNLSKGSVVASYRKTHLDVE	164
Db	95	LMAQYRELAKCKMKIMTSLSGVNER-----NDQKITFAHYHLNMEKGILAYVRKLHNPDVT	149
Oy	165	IPDGCPMRNSNTYTKRGGLTEPPVKTGRACKVGLAICTYDMRFPELSIKLADAGAEILLTPSA	224
Db	150	TK-EVRLRESDDVTPTGYCLESREPVSTPVGIOIGIQCYCDLDFAPPAVLIRKLGANLLTPSA	208
Oy	225	PQSIVGPAHWELLARALESCQYIIAQAOCGRHHETRASVGHSNWVPDGVVAARCSCG	284
Db	209	FYYAVGKRWHELLARALETCQEVVAALIAQGHNOKROSKWSGHSMIVTSFGNVALLDCSEO	268
Oy	285	P-GLCIARIIDLHFLOOMRONLPVPOHRRPDYTGSIHPL 322	
Db	269	EIDDTAEVDLSVLDOSLYQTMPCFEHRRRDITALTAYNL 307	

Q9LE50  
ID Q9LE50 PRELIMINARY; PRT; 316 AA.  
AC Q9LE50;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE NITRILASE 1 LIKE PROTEIN.  
GN ATG08790.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OC NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McCombie R.W., Spiegel L.A., Huang E.N., Nascimento L.U.,  
RA de la Bastide M., Vil D.M., Preston R.R., Matero A., Shah R.,  
RA O'Shaughnessy A., Rodriguez M., Shekher M., Schult K., See L.H.,  
RA Swaby I., Habermann K., Dedhia N.N., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,  
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,  
RA Shekher M., Schult K., See L.H., Swaby I., Habermann K., Dedhia N.N.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL161812; CAB82115.1; -;  
DR EMBL: AL161512; CAB78004.1; -;  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase.1.  
SQ SEQUENCE 316 AA; 34807 MW; 1D740F48DF9A202C CRC64;

Query Match 36.4%; Score 632.5; DB 10; Length 316;  
Best Local Similarity 45.2%; Pred. No. 6.1e-50;  
Matches 126; Conservative 56; Mismatches 88; Indels 9; Gaps 5;

QY 31 PRTMSSTSWELPLVAVCOVTPSPKQENFTCAELVQEAARACACLAFLPEAFPIARN 90  
DB 26 PLTMATTVN-KITRVAAQAQMTSVNDLMTNPTATCSRLVQEAALAGAKLCPPENFSYGDK 84  
QY 91 PAETLLSEPLNGDLLQYQSOLARECGIWLSLGFEHFGODWEONKIYNCHVLLNSKS 150  
DB 85 EGSVKIAEPLDGPVMEYRCSLARSDNIMWLSLGFQERPD---THLCMTHVVIDAGM 140  
QY 151 VVASYKTHLCDEVEIPGQGMRESNNTKPGCTLEPPYKTAGKGLAICGDMPELSLK 210  
DB 141 IRDTYKMHLEFDVDPGSSYSKSSFTVP-GTKIVSDSEVGRIGLVICYDLRFPKITYQ 199  
QY 211 LA-QAGAEILTVSAFGSVTPGAHWEVLRLARAIESCQYIAAAGCRHHESTRASGSHM 269  
DB 200 LRFQQAQVLLVPSATFKVGEAHWEVLRLARAIETQYIAAQAQKHNKESYDGL 259  
QY 270 VDPMGTVVAVRCSE--GPGICLARIDLHFLQOMROHLPV 306  
DB 260 IIDPMGVVGVRLPDRVSTGIIVADIDPSLIDSVRTKMPI 298

RESULT 9  
ID 094660 PRELIMINARY; PRT; 276 AA.  
AC 094660;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HYPOTHEICAL 30.4 KDA NITRILASE-LIKE PROTEIN C651.02 IN CHROMOSOME II.  
GN SPBC651.02.

OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OC NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-972;  
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO THE N-TERMINAL OF NITRILASES.  
DR EMBL: AL035570; CAB37598.1; -;  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase.1.  
KW Hypothetical protein: Lyase.  
KW ACT\_SITE 156 156  
FT ACT\_SITE 156 156  
SQ SEQUENCE 276 AA; 30421 MW; 07AA741A54297E07 CRC64;

Query Match 33.9%; Score 588.5; DB 3; Length 276;  
Best Local Similarity 45.8%; Pred. No. 5.5e-46;  
Matches 132; Conservative 45; Mismatches 88; Indels 23; Gaps 8;

QY 42 LPIVAVCOVTPSPKQENFTCAELVQEAARACACLAFLPEAFPIARNPAETL-LTSEP 100  
DB 1 MTLAAVAQNLSSGSLKLNLAICKELISQAAGAKKCIFFPASPFIHNSDEATLELNHP 60  
QY 101 LNSDGLQYQSOLARECGIWLSLGFEHFGODWEONKIYNCHVLLNSK-----GSVA 153  
DB 61 DCSKFIDVESAPKHSIFVNI-CVHE-----PSKVKN--KLLNSGLFIEPLHGETIS 110  
QY 154 SYRTHLCDEVEIPGQGMRESNNTKPGCTLEPPYKTAGKGLAICGDMPELSLKLAQ 213  
DB 111 RYSKAHLFDVEIKNGPPLTKESNTTLRGEALIPCKTPGKVSACIFDIRPEQATIKRN 170  
QY 214 AGAEILTVPSAFSGSVTPGAHWEVLRLARAIESCQYIAAAGCRHHESTRASGSHMVPD 273  
DB 171 MGAHIIITPSAFTEKTKGAHWEVLRLARALDSQCYIAPAGQGNKRNKRSYGSMTYDP 230  
QY 274 WGVTVAVRCSE--GP-GICLARIDLHFLQOMROHLPVORHRRPYLSL 318  
DB 231 WGVTVAVRCSE--GP-GICLARIDLHFLQOMROHLPVORHRRPYLSL 318

RESULT 10  
ID 027839 PRELIMINARY; PRT; 272 AA.  
AC 027839;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE N-CARBAMOYL-D-AMINO ACID AMIDOHYDROLASE.  
GN MTH1811.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OC Methanothermobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-DELTA H;  
RC MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwanji N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Nollan J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
deltaH: functional analysis and comparative genomics."  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL: AE000934; AAB86277.1; -;  
DR InterPro: IPR003010; CN\_hydrolase.



```

DE      01-JUN-2001 (TREMBLE, 17, Last annotation update)
DN      HYPOTHEETICAL PROTEIN PA4475.
GN      PA4475.
OS      Pseudomonas aeruginosa.
OC      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC      Pseudomonas.
OX      NCBI_TaxID=287;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=PA01;
RX      MEDLINE=20437337; PubMed=10984043;
RX      Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA      Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA      Garber R.L., Goltzy L., Tolentino E., Westbrock-Medman S., Yuan Y.,
RA      Brody L.L., Coulter S.N., Folger K.R., Kas A., Lablidy K., Lam R.M.,
RA      Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA      Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT      "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT      opportunistic pathogen.";
RL      Nature 406:959-964 (2000).
RL      EMBL; AE004861; AAG07863.1; -.
DR      InterPro: IPR003010; CN_hydrolase.
DR      InterPro: IPR001110; UPE0012.
DR      Pfam: PF00795; CN_hydrolase; 1.
DR      PROSITE, PS01227; UPE0012; 1.
DR      Hypothetical protein; Complete proteome.
KW      SEQUENCE 282 AA; 30724 MW; 335ADC799F785584 CRC64;

```

Query Match	27.8%	Score 482	DB 2	Length 282
Best Local Similarity	40.98%	Pred. No. 3.2e36		
Matches 110	Conservative 45	Mismatches 110	Indels 4	Gaps 4

[illegible]

RESULT	14		
09JHM2			
ID	Q9JHM2	PRELIMINARY:	PRT: 276 AA.
AC	Q9JHM2:		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	NIT PROTEIN 2 (1190017B19RIK PROTEIN).		
GN	NIT2 OR 1190017B19RIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Mammalia; Rodentia; Sciurognathi; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	Pubmed=10959838;		
RA	Pace H.C., Hodawadekar S.C., Draganescu A., Huang J., Bieganski P.,		
RA	Pekarsky Y., Croce C.M., Brenner C.;		
RT	"Crystal structure of the worm NitFhlt Rosetta Stone protein reveals a		

RT Nit tetramer binding two Pht dimers." ;  
 RL Curr. Biol. 10:907-917(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aikawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S.,  
 RA Saito T., Okazaki Y., Golboori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batlvo S., Casavant T.,  
 RA Fleischiemann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nakato I., Pesole G., Quackenbush J.,  
 RA Schirrali L.M., Stambli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Grossstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniwa M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,  
 RA Nordone P., Rino B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,  
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 DR EMBL: AF284573; AAF87102.1; -  
 DR EMBL: AK004535; BAB23354.1; -  
 DR MGD: MGI:1913477; 1190017B19Rk.  
 DR InterPro: IPR003010; CN\_hydrolase.  
 SO Pfam: PF00795; CN\_hydrolase; 1.  
 SO SEQUENCE 276 AA; 30501 MW; 1740FDC44978326D6 CRC64;

	Query Match	27.1%	Score 470.5	DB 11	Length 276
	Best Local Similarity	37.6%	Pred. No. 3.5e-35		
	Matches 103	Conservative 50	Mismatches 108	Indels 13	Gaps 4
QY	45	VAVCVCTSPNNKOEHNKCAELVQEAARAGACATLPEAFDFTIARNPAPETLL	----	SEPT	101
DB	6	LALIDLOVSSLSKSDULTRACSLVREDAKAGANIVSLPECFN	----	SPYGTTFPDDYAEKI	61
QY	102	NGDLLGOYLSQALAREGIMVSLSGFPERGQDMQONKIYCHVLLNSKGSVA	SYERTHLIC	161	
DB	62	PEESYQKLESEVAKESSTIYLIGSSIFE	----	EDAKRLYTCVCFPGDGLVKKHKRIHLE	116
QY	162	DVEIDPGQGMRESNNTYKPGTLEPPVKTPAGKVGALICYDMRPPELSIKLAQAGAEILTY		221	
DB	117	DIDVPGKITFGGSKTILSPDSFS	-TFDPFYCKVGGICYDMNFAELQIYARGOGLLVY	175	
QY	222	PSAFGSVTGPRAWEYLLRRARAEISOCYIYAAACQGHHESTRASVYHSDPWGIVAVARC		281	
DB	176	PEAFNLTGTPAHMELJORARAAVDQNVYATASPARDDKASVAMGSHSTVDPWGQVLLTKA		235	
QY	282	SESGPGCLARIDLHFLQOMRQHLPVFGHRRDLY		315	
DB	236	GTEETILVSDIDLKRLAEIRQDIPILKOKRALLY		269	

RESULT	15	
09KU04		
ID	09KU04	PRELIMINARY; PRT: 275 AA.
AC	09KU04:	
DT	01-OCT-2000	(TREMBLrel. 15, Created)
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)
DE	HYPOTHETICAL PROTEIN VC0421.	
GN	VC0421.	
OS	Vibrrio cholerae.	
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.	
OX	NCBI_TaxID=66;	
RN	[1]	
RP	SEQUENCE FROM N.A.	



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: April 26, 2002, 16:55:51 ; Search time 33.36 Seconds

(without alignments)  
737.541 Million cell updates/sec

Title: US-09-357-675C-22

Perfect score: 1736  
Sequence: 1 MGFITRPQGLCTGYRL.....LPVFGHRRPDLXSGIHPHS 323Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summariesDatabase : PIR.68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	706	40.7	440	2	T43198
2	632.5	36.4	316	2	D85088
3	588.5	33.9	276	2	T40601
4	538.5	31.0	272	2	B69109
5	523	30.1	307	2	S56907
6	492	28.3	272	2	S76524
7	482	27.8	282	2	E83086
8	437.5	25.2	322	2	T38399
9	431.5	24.9	275	2	F82325
10	419	24.1	291	2	S51459
11	417	24.0	318	2	T48563
12	410	23.6	270	2	B81199
13	404	23.3	270	2	E81834
14	341.5	19.7	264	2	T36488
15	340	19.6	257	2	C69264
16	334	19.2	297	2	F75263
17	314	18.1	187	2	H64796
18	301	17.3	187	2	D85562
19	285.5	16.4	292	2	G83608
20	276.5	15.9	259	2	E69863
21	276.5	15.9	271	2	G83780
22	275	15.8	287	2	A72454
23	271	15.6	270	2	B83387
24	268.5	15.5	340	2	C70743
25	268	15.4	285	2	S58240
26	262.5	15.1	220	2	A84673
27	262.5	15.1	280	2	T34905
28	262.5	15.1	295	2	H82556
29	255.5	14.7	280	2	T28684

30	255	14.7	298	2	T17568	hydrolase homolog
31	251.5	14.5	292	2	E64558	conserved hypotet
32	250.5	14.4	294	2	G71949	hypothetical prote
33	249.5	14.4	261	2	S20793	hypothetical prote
34	248	14.3	292	2	E64614	beta-alanine synth
35	245	14.1	292	2	F71901	hypothetical prote
36	226.5	13.0	262	2	C71109	hypothetical prote
37	225.5	13.0	271	2	H83195	conserved hypotet
38	225	13.0	272	2	T41662	probable nitrilase
39	222	12.8	262	2	C75051	hypotet related
40	222	12.8	290	2	B81369	probable hydrolase
41	214.5	12.4	246	2	A70310	conserved hypotet
42	208.5	12.0	346	2	T49147	nitrilase (EC 3.5.
43	206	11.9	346	2	S22398	nitrilase (EC 3.5.
44	203	11.7	346	2	T49148	nitrilase (EC 3.5.
45	199	11.5	304	2	JW0082	N-carbamyl-D-amino

## ALIGNMENTS

```
RESULT 1
T43198
nitrilase/Fhit protein - Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence=revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43198
R:pekarsky, Y.; Campiglio, M.; Siprashvili, Z.; Druck, T.; Sedkov, Y.; Tiliib, S.; Dr
Proc. Natl. Acad. Sci. U.S.A. 95, 8744-8749, 1998
A:Title: Nitrilase and Fhit homologs are encoded as fusion proteins in Drosophila mel
A:Reference number: Z22332; MUID:98337986
A:Accession: T43198
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-440 <PEP>
A:Cross-references: EMBL:AF069986; NID:g3228663; PIDN:AMC39136.1; PID:g3228664
C:Genetics:
A:Gene: NITPHIT

Query Match 40.7%; Score 706; DB 2; Length 440;
Best local similarity 48.6%; Pred. No. 7.1e-56;
Matches 138; Conservative 51; Mismatches 89; Indels 6; Gaps 2;

OY 32 RIMSSSTWELPLVAVCOVTSIPNKOENKTCALVQEAARLGAFLPEAFDFTARNP 91
|||: : |||||: : |||: : |||||: :
DB 8 RFWATGRHF---IAVCOMTSNDLDEKNFOAAKNMTERAGEKKCEWVFLPECFDFTGLNK 63
|||: : |||||: : |||: : |||||: :
OY 92 AETLLSEPLNGDLLGOYSOLARECGIWTSLGFEHFGQDWEQONKIYNCHVLLNKGVS 151
||: : : : : : |||||: : |||||: : |||||: : |||||: : |||||: :
DB 64 NEQIDLAMATDCMYEMKYRELARKHNWLSLGLHKK--DPSAHPWNTHLITDSGYT 121
||: : : : : : |||||: : |||||: : |||||: : |||||: : |||||: :
OY 152 VASRYRTHLCDVEIPGCGPMBRESNYTRPGCTLEPPVKTPAGKGLAICYDMRPETSLKL 211
|||: : |||||: : |||: : |||||: : |||||: : |||||: : |||||: :
DB 122 RAYNNLHLEFDELPKRVLRMESEFSKAGTEMIPVDFTIGRGLSLICYDVRPELSTWN 181
||: : : : : : |||||: : |||||: : |||||: : |||||: : |||||: :
OY 212 AQAQAEILTPPSAFVSGVTPGAHWEVLLRARAIISQCYVIAAOCGRHHETRASVGHSMV 271
||: : : : : : |||||: : |||||: : |||||: : |||||: : |||||: :
DB 182 RKGGAQLLSTFSPSAFTLNTGLAHWETLLRARAIENOCYVAAAOAGTGAHNPFGSGHSMV 241
||: : : : : : |||||: : |||||: : |||||: : |||||: : |||||: :
OY 272 DPMGAVVACSESGPGLARIDHFLQOMRONHLPVQHRPDLX 315
||| ||||| ||||| : ||||| : : ||||| ||||| ||||| ||||| |||||
DB 242 DPMGAVVACSESRVDMCFEIDLSTVDLTLEKMPVESHRSRLY 285
||| ||||| ||||| : ||||| : : ||||| ||||| ||||| ||||| |||||

RESULT 2
nitrilase 1 like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold sp
Nature 402, 769-777, 1999
```









A: Experimental source: cultivar Columbia; BAC clone F14F18  
 C: Genetics:  
 A: Map position: 5  
 A: Introns: 25/3; 64/3; 139/3; 182/1; 207/1; 233/2; 238/3; 265/3; 293/2  
 A: Note: F14F18.210

Query Match 24.0%; Score 417; DB 2; Length 318;  
 Best Local Similarity 33.2%; Pred. No. 5.7e-30;  
 Matches 105; Conservative 48; Mismatches 123; Indels 40; Gaps 8;

21 RIPVLCTOPRPRMTSSSTWELPL---VAVCVSTPNKQENFKTCAELVQEARLGAC 76  
 11 RVPNALPLPAP-----PLTKFNIGLCQLSVTSQKRNRIISAKKAIEEAASGAK 59  
 77 LAFLEPAFPIARNAETLLISE-PLNGDL---LGOYSOLARECGIWLISGHERGQDM 132  
 60 LVLLPEIMSPYSNDSPFYAEIEDAGDASPSTAMLSVSKRLKTIIGSISPERVGD- 118  
 133 EQNOKIYNCHVLLNSKGSVVASRKTHLCDVEIPGCGPRRESNVTKPGGTLE----- 184  
 119 ---RKYNCVYSGDGEELAKHRIKTHLPIDIPGKITMESKTTLAGETPTIVDTGYML 174  
 185 ---PVYKTPAGKYLAIICYDMRPELSLKLQAQAEILTPSAFGSVTPGAHWEVLRAA 242  
 175 GLPNIIPDGRIGIGICVDIRFOLAMITYAARGAHLICYPGAFNMTTGRLHMLLORARA 234  
 243 ISSQCYVIAAOCGRHNETRASY--GHSVVDPWGTGVVAVARCSGPGICLARIIDHFLQO 299  
 235 TTONLVV---ATCSPARDGAGYTAGHSTLVGPFGEVLATHEHEBAITIAIEDYSILRQ 291  
 300 MRQHLVPVQHRPRDLY 315  
 292 RRTSLPLNRQRGDL 307

RESULT 12  
 B81199

nitrlase NMB0441 [imported] - Neisseria meningitidis (strain MC58 serogroup B)  
 C: Species: Neisseria meningitidis  
 C: Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C: Accession: B81199  
 R: Tettein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischman, R.D.; Dougherty, B.A.;  
 Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Mastignani, V.; Pizza, M.  
 Science 287, 1809-1815, 2000  
 A: Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
 A: Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A: Reference number: A81000; MUID:20175755  
 A: Accession: B81199  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-270 <TEXT>  
 A: Cross-references: GB:AE002400; GB:AE002098; NID:g7225659; PIDN:AAF40879.1; PID:g722566  
 A: Experimental source: serogroup B, strain MC58  
 C: Genetics:  
 A: Gene: NMB0441  
 C: Superfamily: hypothetical protein YLR351c

Query Match 23.6%; Score 410; DB 2; Length 270;  
 Best Local Similarity 37.5%; Pred. No. 2e-29;  
 Matches 103; Conservative 39; Mismatches 109; Indels 24; Gaps 8;

45 VAVCVSTPNKQENFKTCAELVQEARLGACIAFLPEAFDFIARNAETLLISEPLNGD 104  
 6 VAAVQVNSVSPETNVAAAMKRLVARAAEGADVLLPDEVVLMGANDTDLALAEPLGG- 64  
 105 LIGQY---SOLARECGIWLISG-----FHERGQDMQONOKIYNCHVLLNSKGSVVAS 155  
 65 --GRFQALSEIATKEGGVVL-FGGTVPLQSCENG-----KVMNTLLVYGRDGVRTGLY 114  
 156 RKTTHLCDVEIPGCGPRRESNVTKPGGTLEPRPVYKTPAGKYLAIICYDMRPELSLKLQAG 215

115 HKMHLFGFSGIGLGE-RYAEADTIRAGDV--PHISAEGVPAAGICVDYRPFEPFRR--QLP 170  
 216 AELLTPSAFGSVTPGAHWEVLRLARAISSQCYVIAAOCGRHNETRASYGHSVVDPWG 275  
 171 FVILMPAAFTHTTGKAHWEVLLRARAVENOCYVVAAGGGLHENGRRFGHSMIVDPWG 230  
 276 TVVARCSGPGICLARIIDHFLQDMRQHLVPVQHR 310  
 231 DVDVLPREGGVYTAIDANRLNSVRNRLPALKYR 265

RESULT 13  
 E81834

conserved hypothetical protein NMA2044 [imported] - Neisseria meningitidis (strain Z2  
 C: Species: Neisseria meningitidis  
 C: Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C: Accession: E81834  
 R: Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
 Nature 404, 502-506, 2000  
 A: Reference number: A81775; MUID:20222556  
 A: Accession: E81834  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-270 <PAR>  
 A: Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85262.1; PID:g738  
 A: Experimental source: serogroup A, strain Z2491  
 C: Genetics:  
 A: Gene: opaA; NMA2044  
 C: Superfamily: hypothetical protein YLR351c

Query Match 23.3%; Score 404; DB 2; Length 270;  
 Best Local Similarity 37.1%; Pred. No. 6.8e-29;  
 Matches 102; Conservative 34; Mismatches 115; Indels 24; Gaps 7;

45 VAVCVSTPNKQENFKTCAELVQEARLGACIAFLPEAFDFIARNAETLLISEPLNGD 104  
 6 VAAVQVNSVSPETNVAAAMKRLVARAAEGADVLLPDEVVLMGANDTDLALAEPLGG- 64  
 105 LIGQY---SOLARECGIWLISG-----FHERGQDMQONOKIYNCHVLLNSKGSVVAS 155  
 65 --GRFQALSEIATKEGGVVL-FGGTVPLQSPENG-----KVMNTLLVYGRDGRKRTGLY 114  
 156 RKTTHLCDVEIPGCGPRRESNVTKPGGTLEPRPVYKTPAGKYLAIICYDMRPELSLKLQAG 215  
 115 HKMHL--FGFSGLGELYAEADTILAGREVRHLLAEGSVAAAGICVDYRPFEPFRR--QLP 170  
 216 AELLTPSAFGSVTPGAHWEVLRLARAISSQCYVIAAOCGRHNETRASYGHSVVDPWG 275  
 171 FVILMPAAFTHTTGKAHWEVLLRARAVENOCYVVAAGGGLHENGRRFGHSMIVDPWG 230  
 276 TVVARCSGPGICLARIIDHFLQDMRQHLVPVQHR 310  
 231 DVDVLPREGGVYTAIDANRLNSVRNRLPALKYR 265

RESULT 14  
 T36488

probable hydrolase - Streptomyces coelicolor  
 C: Species: Streptomyces coelicolor  
 C: Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C: Accession: T36488  
 R: Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M  
 submitted to the EMBL Data Library, July 1999  
 A: Reference number: Z21608  
 A: Accession: T36488  
 A: Status: preliminary; translated from GB/EMBL/DBJ  
 A: Molecule type: DNA  
 A: Residues: 1-264 <SAU>  
 A: Cross-references: EMBL:AL096822; PIDN:CAB46930.1; GSPDB:GN00070; SCOEDB:SCGD3.08C





THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:54:36 ; Search time 56.94 seconds  
(without alignments)  
420.191 Million cell updates/sec

Title: US-09-357-675c-22  
Perfect score: 1736  
Sequence: 1 MLCFTRPPHQLCTGYRL.....LPVQRHRPDLVSGHPLS 323

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
A.Geneseq\_1101.\*  
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1482	85.4	464	21	AAV68739
2	747	43.0	224	21	AAV68739
3	733	42.2	153	20	AAV10877
4	483	27.8	276	22	AAAB80984
5	313	18.0	159	21	AAAB42967
6	295.5	17.0	266	22	AAAG2814
7	288.5	16.6	299	21	AAAG09683
8	288.5	16.6	299	21	AAAG4945
9	273.5	15.8	252	22	AAAB79700
10	273.5	15.8	252	22	AAAB80193
11	262.5	15.1	220	21	AAAG09684

12	262.5	15.1	220	21	AAAG49946	Arabidopsis thaliana
13	237	13.7	264	22	AAAG1841	S. epidermidis ope
14	210	12.1	227	18	AAV55670	H. pylori ORF hp5p
15	205	11.8	340	21	AAAG44265	Arabidopsis thaliana
16	205	11.8	346	21	AAAG44265	Arabidopsis thaliana
17	205	11.8	358	21	AAAG44264	Arabidopsis thaliana
18	204	11.7	303	15	AAAR46251	Heat-stable carban
19	203	11.7	303	15	AAAR46258	Heat-stable carban
20	202	11.6	303	15	AAAR46241	Heat-stable carban
21	201.5	11.6	303	15	AAAR46260	Heat-stable carban
22	201	11.5	303	15	AAAR46259	Heat-stable carban
23	200.5	11.5	303	15	AAAR46268	Heat-stable carban
24	200	11.5	303	15	AAAR46249	Heat-stable carban
25	200	11.5	303	15	AAAR46243	Heat-stable carban
26	200	11.5	303	15	AAAR46242	Heat-stable carban
27	200	11.5	303	15	AAAR46244	Heat-stable carban
28	200	11.5	303	15	AAAR46265	Heat-stable carban
29	200	11.5	303	15	AAAR46272	Heat-stable carban
30	199.5	11.5	303	15	AAAR46261	Heat-stable carban
31	199.5	11.5	303	15	AAAR46269	Heat-stable carban
32	199.5	11.5	303	15	AAAR46270	Heat-stable carban
33	199	11.5	303	15	AAAR46252	Heat-stable carban
34	199	11.5	303	15	AAAR46253	Heat-stable carban
35	199	11.5	303	15	AAAR46250	Heat-stable carban
36	199	11.5	303	15	AAAR46239	Heat-stable carban
37	199	11.5	303	15	AAAR46245	Heat-stable carban
38	199	11.5	303	15	AAAR46246	Heat-stable carban
39	199	11.5	303	15	AAAR46264	Heat-stable carban
40	199	11.5	303	22	AAAB74624	Agrobacterium sp.
41	198.5	11.4	303	15	AAAR46257	Improved Heat-stab
42	198.5	11.4	303	15	AAAR46255	Improved Heat-stab
43	198.5	11.4	303	15	AAAR46256	Improved Heat-stab
44	198.5	11.4	303	15	AAAR46267	Improved Heat-stab
45	198.5	11.4	303	22	AAAB74625	Escherichia coli d

## ALIGNMENTS

RESULT 1	
AAV68739	standard; Protein; 464 AA.
ID	AAV68739
AC	AAV68739;
DT	05-MAY-2000 (first entry)
DE	Amino acid sequence of the N1L protein.
XX	
KW	NIIL gene; nitrate; tumour suppressor gene; FHIT; chromosome 3p14.2;
KW	FRAB; cancer; genome allele inactivation.
XX	
OS	Homo sapiens.
OS	Mus sp.
OS	Drosophila melanogaster.
OS	Caenorhabditis elegans.
XX	
FT	Key
FT	Misc-difference 6
FT	Location/Qualifiers
FT	/note="unspecified amino acid encoded by TNT"
PN	WO200003685-A2.
XX	
PD	27-JAN-2000.
XX	
PF	20-JUL-1999; 99WO-US16366.
XX	
PR	20-JUL-1998; 98US-0093350.
XX	
PA	(UYUE-) UNIV JEFFERSON THOMAS.
XX	
PI	Croce CM;
XX	





prostate disease; asthma; osteoporosis; arthritis.  
 Homo sapiens.  
 Key Location/Qualifiers  
 Peptide 1..46  
 FT /note= "signal peptide"  
 FT 47..153  
 FT Protein /note= "secreted protein"  
 M09907891-A1.  
 18-FEB-1999.  
 04-AUG-1998; 98WO-US16235.  
 19-AUG-1997; 97US-0056732.  
 PR 05-AUG-1997; 97US-0054798.  
 PR 05-AUG-1997; 97US-0054803.  
 PR 05-AUG-1997; 97US-0054804.  
 PR 05-AUG-1997; 97US-0054806.  
 PR 05-AUG-1997; 97US-0054807.  
 PR 05-AUG-1997; 97US-0054808.  
 PR 05-AUG-1997; 97US-0054809.  
 PR 05-AUG-1997; 97US-0055309.  
 PR 05-AUG-1997; 97US-0055310.  
 PR 05-AUG-1997; 97US-0055312.  
 PR 05-AUG-1997; 97US-0055386.  
 PR 05-AUG-1997; 97US-0055311.  
 PR 18-AUG-1997; 97US-0055970.  
 PR 18-AUG-1997; 97US-0055986.  
 PR 19-AUG-1997; 97US-0056365.  
 PR 19-AUG-1997; 97US-0056366.  
 PR 19-AUG-1997; 97US-0056367.  
 PR 19-AUG-1997; 97US-0056370.  
 PR 19-AUG-1997; 97US-0056371.  
 PR 19-AUG-1997; 97US-0056563.  
 PR 19-AUG-1997; 97US-0056731.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;  
 Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;  
 WPI: 1999-167452/14.  
 N-PSDB: AAX30398.  
 New isolated human genes encoding secreted polypeptides - useful for  
 diagnosis and treatment of pathological diseases  
 Claim 3: Page 304; 331pp; English.  
 The specification describes secreted proteins and their corresponding  
 polynucleotides which are useful for preventing, treating or ameliorating  
 medical conditions, e.g. by protein or gene therapy. Pathological  
 conditions can also be diagnosed by determining the amount of the  
 secreted polypeptides in a sample or by determining the presence of  
 mutations in the polynucleotides. Specific uses are described for each  
 of the products, based on which tissues they are most highly  
 expressed in, and include developing products for the diagnosis or  
 treatment of cancer, tumours, neurodegenerative disorders, developmental  
 abnormalities and foetal deficiencies, blood disorders, CNS disorders,  
 diseases of the immune system, autoimmune diseases, hepatic and renal  
 diseases, diabetes, inflammation, allergies, ischemic shock, Alzheimer's  
 and cognitive disorders, schizophrenia, cardiovascular disorders,  
 prostate diseases, asthma, disorders involving osteoclasts such as  
 osteoporosis, arthritis or malignancies, diseases of testes, lung or  
 thymus, digestive/endocrine disorders, infections and AIDS. The  
 polypeptides are also useful for identifying their binding partners.

Query Match 42.2%; Score 733; DB 20; Length 153;  
 Best Local Similarity 86.9%; Pred. No. 2.3e-68;  
 Matches 133; Conservative 12; Mismatches 8; Indels 0; Gaps 0;  
 QY 171 MRESNYTKPGGTLPEPVKTPAGKVGAIICYDMRPPLSLKLAQAGELITYPSAFSGSVTG 230  
 Db 1 mresnltmpgslpsvscpdkiglavcydmrpfelslalaqagaelitypsafsgitg 60  
 QY 231 PAHWEVLLRARALESCCYIAAAGCGRHHETRASGHSWVPDWCIVVAVRCSPGLCTA 290  
 Db 61 pahwevllraralelscqyvaaagcgrhhekrasysghsmvdpwclvavrcsepglclta 120  
 QY 291 RIDLHFLQOMROHLPVFGHRPDLVSLGHPIS 323  
 Db 121 ridlhyrlqrhlhlpvfqhrpdlvsgnlghpls 153  
 RESULT 4  
 AAB80984  
 ID AAB80984 standard; Protein; 276 AA.  
 XX AAB80984;  
 AC AAB80984;  
 DT 12-JUN-2001 (first entry)  
 XX  
 DE Human hnit3-ase.  
 XX  
 KW Human; hnit3-ase; Cushings adrenal gland tumour;  
 KW nitrile hydrolytic enzyme; nitrilase.  
 XX  
 OS Homo sapiens.  
 XX  
 PN CN1277998-A.  
 XX  
 PD 27-DEC-2000.  
 XX  
 PF 30-MAY-2000; 2000CN-0116221.  
 XX  
 PR 30-MAY-2000; 2000CN-0116221.  
 XX  
 PA (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.  
 XX  
 PI Xu X, Qian B, Zhang X;  
 XX  
 DR WPI: 2001-245678/26.  
 DR N-PSDB: AAF84214.  
 XX  
 PT Human nitrile hydrolytic enzyme protein and its coding sequence -  
 PS Claim 2: Page 11; 20pp; Chinese.  
 XX  
 CC The present invention relates to human hnit3-ase (nitrilase) protein,  
 CC which is expressed in Cushings adrenal gland tumour, and its coding  
 CC sequence (AAF84214 and AAB80984). The present invention also relates to a  
 CC preparation method of said protein and nucleic acid sequence, and a  
 CC method of detecting human hnit3-ase nucleic acid sequence and polypeptide  
 CC in sample.  
 XX  
 SQ Sequence 276 AA;  
 Query Match 27.8%; Score 483; DB 22; Length 276;  
 Best Local Similarity 38.2%; Pred. No. 6.7e-42;  
 Matches 107; Conservative 54; Mismatches 105; Indels 14; Gaps 6;  
 QY 38 TSEELPLVAVCVTSNPNOENFKTCALVQEARLGAACIAPFPAFD--FIARNPAETL 95  
 Db 2 tselrlalql-qtsst--ksdnvtracsfireaatgakiivsipectfnspygakylfpe-- 56  
 QY 96 LISEPLNGDLIGQYSOLARCGIWLISGFHERGQDMEONOKITYNCHVLNLSKGSVAVSY 155  
 Db 57 -yeaklpgestqlksevakecsilylgsispe-----edagklyncavtfgpdgtllaky 110



CC The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 266 AA;

Query Match 17.0%; Score 295.5; DB 22; Length 266;  
Best Local Similarity 32.1%; Pred. No. 2.3e-22;  
Matches 89; Conservative 46; Mismatches 113; Indels 29; Gaps 10;

QY 45 VAVCVSTPKNQENFKCAELVQEARLGAFLPER-PFFIARNPARETLLLEPLNG 103

Db 3 Iallqistnsdmdfiallrdaekaaeggarvlvfpfeatsqsfygrldt--qgeeldg 60

QY 104 DLGQYSLARECGIMLSLGFH-----ERGDWENONKRYNCHVILNKGSGVASYRKT 158

Db 61 efstavrkladeidvviagmftpadvtgrg---ektisrvmtvlisag-ihgynyki 116

QY 159 HLCQVEIDFGQGBRESNNTKPGGTLEPPYKTPAGVGLAICYDMRPPELSLKLQAGAEI 218

Db 117 htyda-----fygredsvkpgdel-vvfevdikfyvatcydtrfpegfkdlarngaqi 170

QY 219 LTYPSAFSGSVGPA---HWEVILLRARELSQCYVIAAOCGNHETR-----ASTGSHM 269

Db 171 Ivpvpswdg--gpgkleqewvlprratdstcwiavcgqarlpeelrderkypgtigsham 228

QY 270 VWDPMGVAVARCEGPGCLARIDLFHQOMRHLPV 306

Db 229 vtnphgeviassgvepmliadiavsglakirealpv 265

RESULT 7

ID AAC09683 standard; Protein: 299 AA.

AC AAC09683;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 7710.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999;

PR 19-APR-1999;

PR 21-APR-1999;

PR 23-APR-1999;

PR 23-APR-1999;

PR 28-APR-1999;

PR 30-APR-1999;

PR 04-MAY-1999;

PR 05-MAY-1999;

PR 06-MAY-1999;

PR 07-MAY-1999;

PR 11-MAY-1999;

PR 14-MAY-1999;

PR 14-MAY-1999;

PR 14-MAY-1999;

PR 18-MAY-1999;

PR 19-MAY-1999;

PR 20-MAY-1999;

PR 21-MAY-1999;

PR 24-MAY-1999;

PR 25-MAY-1999;

PR 27-MAY-1999;

PR 28-MAY-1999;

PR 03-JUN-1999;

PR 04-JUN-1999;

PR 07-JUN-1999;

PR 08-JUN-1999;

PR 10-JUN-1999;

PR 14-JUN-1999;

PR 16-JUN-1999;

PR 17-JUN-1999;

PR 18-JUN-1999;

PR 18-JUN-1999;

PR 18-JUN-1999;

PR 18-JUN-1999;

PR 18-JUN-1999;

99US-0129845.

99US-0130077.

99US-0130449.

99US-0130510.

99US-0130891.

99US-0131449.

99US-0132048.

99US-0132407.

99US-0132484.

99US-0132485.

99US-0132486.

99US-0132487.

99US-0132863.

99US-0134218.

99US-0134219.

99US-0134221.

99US-0134370.

99US-0134768.

99US-0134941.

99US-0135124.

99US-0135353.

99US-0135629.

99US-0136021.

99US-0136392.

99US-0136782.

99US-0137222.

99US-0137528.

99US-0137502.

99US-0137724.

99US-0138094.

99US-0138540.

99US-0138847.

99US-0138119.

99US-0139452.

99US-0139453.

99US-0139492.

99US-0139454.

99US-0139455.

99US-0139456.

99US-0139457.

99US-0139458.

99US-0139459.

99US-0139460.

99US-0139461.

99US-0139462.

99US-0139463.

99US-0139750.

99US-0139763.

99US-0139817.

99US-0139899.

99US-0140353.

99US-0140354.

99US-0140695.

99US-0140823.

99US-0140991.

99US-0141287.

99US-0141842.

99US-0142154.

99US-0142055.

99US-0142390.

99US-0142803.

99US-0142920.

99US-0142977.

99US-0143542.

99US-0143624.

99US-0144005.

99US-0144085.

99US-0144086.

99US-0144325.

99US-0144331.

99US-0144332.

99US-0144333.

PR	19-JUL-1999;	99US-0144334	PR	14-OCT-1999;	99US-0159329
PR	19-JUL-1999;	99US-0144335	PR	14-OCT-1999;	99US-0159330
PR	20-JUL-1999;	99US-0144352	PR	14-OCT-1999;	99US-0159331
PR	20-JUL-1999;	99US-0144353	PR	14-OCT-1999;	99US-0159337
PR	20-JUL-1999;	99US-0144632	PR	14-OCT-1999;	99US-0159637
PR	20-JUL-1999;	99US-0144684	PR	14-OCT-1999;	99US-0159638
PR	21-JUL-1999;	99US-0144814	PR	18-OCT-1999;	99US-0159584
PR	21-JUL-1999;	99US-0145086	PR	21-OCT-1999;	99US-0160741
PR	21-JUL-1999;	99US-0145088	PR	21-OCT-1999;	99US-0160767
PR	22-JUL-1999;	99US-0145085	PR	21-OCT-1999;	99US-0160778
PR	22-JUL-1999;	99US-0145087	PR	21-OCT-1999;	99US-0160770
PR	22-JUL-1999;	99US-0145089	PR	21-OCT-1999;	99US-0160814
PR	22-JUL-1999;	99US-0145192	PR	21-OCT-1999;	99US-0160815
PR	23-JUL-1999;	99US-0145145	PR	22-OCT-1999;	99US-0160980
PR	23-JUL-1999;	99US-0145218	PR	22-OCT-1999;	99US-0160981
PR	23-JUL-1999;	99US-0145324	PR	22-OCT-1999;	99US-0160989
PR	26-JUL-1999;	99US-0145376	PR	25-OCT-1999;	99US-0161404
PR	27-JUL-1999;	99US-0145913	PR	25-OCT-1999;	99US-0161405
PR	27-JUL-1999;	99US-0145918	PR	25-OCT-1999;	99US-0161406
PR	27-JUL-1999;	99US-0145919	PR	26-OCT-1999;	99US-0161359
PR	28-JUL-1999;	99US-0145951	PR	26-OCT-1999;	99US-0161360
PR	02-AUG-1999;	99US-0146386	PR	26-OCT-1999;	99US-0161361
PR	02-AUG-1999;	99US-0146388	PR	28-OCT-1999;	99US-0161920
PR	02-AUG-1999;	99US-0146389	PR	28-OCT-1999;	99US-0161992
PR	03-AUG-1999;	99US-0147038	PR	28-OCT-1999;	99US-0161993
PR	04-AUG-1999;	99US-0147204	PR	29-OCT-1999;	99US-0162142
PR	04-AUG-1999;	99US-0147302			
PR	05-AUG-1999;	99US-0147192			
PR	05-AUG-1999;	99US-0147260			
PR	06-AUG-1999;	99US-0147303			
PR	06-AUG-1999;	99US-0147416			
PR	09-AUG-1999;	99US-0147493			
PR	09-AUG-1999;	99US-0147935			
PR	10-AUG-1999;	99US-0148171			
PR	11-AUG-1999;	99US-0148319			
PR	12-AUG-1999;	99US-0148341			
PR	13-AUG-1999;	99US-0148365			
PR	13-AUG-1999;	99US-0148684			
PR	16-AUG-1999;	99US-0149368			
PR	17-AUG-1999;	99US-0149175			
PR	18-AUG-1999;	99US-0149426			
PR	20-AUG-1999;	99US-0149722			
PR	20-AUG-1999;	99US-0149723			
PR	20-AUG-1999;	99US-0149929			
PR	23-AUG-1999;	99US-0149902			
PR	23-AUG-1999;	99US-0150566			
PR	25-AUG-1999;	99US-0150566			
PR	26-AUG-1999;	99US-0150884			
PR	27-AUG-1999;	99US-0151065			
PR	27-AUG-1999;	99US-0151066			
PR	27-AUG-1999;	99US-0151066			
PR	30-AUG-1999;	99US-0151303			
PR	31-AUG-1999;	99US-0151438			
PR	01-SEP-1999;	99US-0151930			
PR	07-SEP-1999;	99US-0152363			
PR	10-SEP-1999;	99US-0153070			
PR	13-SEP-1999;	99US-0153758			
PR	15-SEP-1999;	99US-0154019			
PR	16-SEP-1999;	99US-0154039			
PR	20-SEP-1999;	99US-0154779			
PR	22-SEP-1999;	99US-0155139			
PR	23-SEP-1999;	99US-0155486			
PR	24-SEP-1999;	99US-0155659			
PR	28-SEP-1999;	99US-0156458			
PR	29-SEP-1999;	99US-0156596			
PR	04-OCT-1999;	99US-0157117			
PR	05-OCT-1999;	99US-0157753			

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132866.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153708.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 16.6%; Score 288.5; DB 21; Length 299;  
Best Local Similarity 28.2%; Pred. No. 1.5e-21;  
Matches 82; Conservative 56; Mismatches 122; Indels 31; Gaps 9;

QY 44 LVACQVSTPNKQNEFTCAELVQEARLGCLAFLEPAFP---FIRNPAETLLSLP 100  
Db 11 vvsslqfscsdistcnuvaaeerlveaahkaganlllqelltegyyfcqagredffktrakp 70  
QY 101 L-NGDLDQVYSOLARECCIMWLSLGFHERGQDWEONKRYNCHVLSKGSVVAASYRKT 159  
Db 71 yknbptlarmqklakelvgvlpvsffe-----eantahynsialldedgtdlgjyrksh 124  
QY 160 LCDVEIPQOGR-MRESNTRKPEGTLIEPPVKTPAGKVGALICYDMRFPPELSLKLQAQAEI 218  
Db 125 lpd-----gpqygckfyfnpqdtgfkvfqtkfaklgyvalcswdqvfpcaaramvlgaei 178  
QY 219 LTVPSAFGS-----VTPRAHWEVLLRARATESQCVVIAAOCGR-----HHERRAS-Y 265  
Db 179 lfypaaissepqdgldrdhmrwymghaganvvpivasnrlgkellelletgspqdlty 238  
QY 266 GHSWVDPWGTAVARCSG-GRGLCLARIDLHFLQOMRQHLPEVQHRPPDLX 315  
Db 239 gtsfiagptglveaeaddkseavlvagfdldmikskrtgswgyfirtrpdlly 289

RESULT 9  
AAB79700  
ID AAB79700 standard; Protein: 252 AA.  
XX  
AC AAB79700;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:134.  
XX  
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.  
XX  
OS Corynebacterium glutamicum.  
XX  
EN W0200100843-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-IB00923.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 14-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032926.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1033004.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 12-AUG-1999; 99US-0148613.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040832.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041380.  
PR 31-AUG-1999; 99DE-1041394.  
PR 31-AUG-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042124.  
PR 03-SEP-1999; 99DE-1042129.  
PR 09-MAR-2000; 2000US-0187970.  
XX  
RA (BADI ) BASF AG.  
XX  
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;









PR 23-APR-1999; 9905-0130510.  
PR 23-APR-1999; 9905-0130891.  
PR 28-APR-1999; 9905-0131449.  
PR 30-APR-1999; 9905-0132007.  
PR 30-APR-1999; 9905-0132407.  
PR 04-MAY-1999; 9905-0132484.  
PR 05-MAY-1999; 9905-0132485.  
PR 06-MAY-1999; 9905-0132486.  
PR 07-MAY-1999; 9905-0132487.  
PR 07-MAY-1999; 9905-0132863.  
PR 11-MAY-1999; 9905-0134256.  
PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134221.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.

PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 28-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149375.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.





[illegible]

PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039
PR	20-SEP-1999;	9905-0154779
PR	22-SEP-1999;	9905-0155131
PR	23-SEP-1999;	9905-0155486
PR	24-SEP-1999;	9905-0155665
PR	28-SEP-1999;	9905-0156458
PR	29-SEP-1999;	9905-0157116
PR	04-OCT-1999;	9905-0157593
PR	05-OCT-1999;	9905-0157753
PR	06-OCT-1999;	9905-0157865
PR	07-OCT-1999;	9905-0158029
PR	08-OCT-1999;	9905-0158232
PR	12-OCT-1999;	9905-0158369
PR	13-OCT-1999;	9905-0158529
PR	13-OCT-1999;	9905-0159294
PR	14-OCT-1999;	9905-0159325
PR	14-OCT-1999;	9905-0159331
PR	14-OCT-1999;	9905-0159637
PR	14-OCT-1999;	9905-0159638
PR	18-OCT-1999;	9905-0159584
PR	21-OCT-1999;	9905-0160741
PR	21-OCT-1999;	9905-0160767
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160770
PR	21-OCT-1999;	9905-0160814
PR	21-OCT-1999;	9905-0160815
PR	22-OCT-1999;	9905-0160980
PR	22-OCT-1999;	9905-0160989
PR	22-OCT-1999;	9905-0161404
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161359
PR	26-OCT-1999;	9905-0161369
PR	26-OCT-1999;	9905-0161361
PR	26-OCT-1999;	9905-0161992
PR	28-OCT-1999;	9905-0161992
PR	29-OCT-1999;	9905-0162142
PR	29-OCT-1999;	9905-0162143

Query Match	11.8%;	Score 205;	DB 21;	Length 340;
Best Local Similarity	24.7%;	Pred. No. 9, 2e-13;		
Matches 82; Conservative	50;	Mismatches 116;	Indels 84;	Gaps 13

```

0Y 30 PRPTMSSTMEELRVANVCQYTSIPNKOENKTKCAELVQEARIGALAPLPEAF----- 84
Db 20 Ivtlvqsvstlvyn-----dpralcbaekf-----iveaskyalvlvlfpeeflggyp 66
0Y 85 -----DFTARNPAETLLISEPLNDLLGOYSQALARECCGWLSSGFGH 126
Db 67 rgffgtglavghneegridef--rnyhasal--kvpppxverlaaelagkmvnhlvmapate 120
0Y 127 EBGDOOMEONKQITYNCHVLLNLSKGSVVASYRKTHLCADVE--IPGGPMRRESNYTRPGGTLE 184
Db 122 kog-----yllyctaifffspqgfgikhrkvmptstlerclwgg-----dgstl 165
0Y 185 PVRVTPACKVGLATCYDMRRPELSLKLAQACAEILITYPASFGSVTGAHMEVLLRARAE 244
Db 166 pyudtrpikgkigaalocwemrmlpyltalyakjleiscaptadysl-----ewqasimihave 220
0Y 245 SQCYVIAAAG--CGR-----NHETASYGHSWVVDPMGCVIARCS--EG 284
Db 222 ggcctvisahqicctkrfeprehodylfndatvdckehorpcvsggsvyisprkivlaagrpves 201
0Y 285 PGLCTARIDLHFLDOOMRONLRFVGH--RRPDLX 315
Db 282 eglvatradidldgiarakylgtavghyskprdlf 313

```

Search completed: April 26, 2002, 16:54:36  
Job time: 234 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:55:11 ; Search time 28.88 Seconds  
(without alignments)  
251.682 Million cell updates/sec

Title: US-09-357-675C-22

Perfect score: 1736

Sequence: 1 MGFTRPPHQLCTGYRL.....LPVQHRRPDLYSGHPLS 323

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents,AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	204	11.8	303	2	US-08-294-871A-26	Sequence 26, Appl
2	204	11.8	303	2	US-08-294-871A-64	Sequence 64, Appl
3	204	11.8	303	3	US-08-876-398A-26	Sequence 26, Appl
4	204	11.8	303	3	US-08-876-398A-64	Sequence 64, Appl
5	203	11.7	303	2	US-08-294-871A-40	Sequence 40, Appl
6	203	11.7	303	2	US-08-876-398A-40	Sequence 40, Appl
7	202	11.6	303	2	US-08-294-871A-6	Sequence 6, Appl
8	202	11.6	303	2	US-08-876-398A-6	Sequence 6, Appl
9	201.5	11.6	303	2	US-08-294-871A-44	Sequence 44, Appl
10	201.5	11.6	303	2	US-08-876-398A-44	Sequence 44, Appl
11	201	11.6	303	2	US-08-294-871A-42	Sequence 42, Appl
12	201	11.6	303	2	US-08-294-871A-62	Sequence 62, Appl
13	201	11.6	303	3	US-08-876-398A-42	Sequence 42, Appl
14	201	11.6	303	3	US-08-876-398A-62	Sequence 62, Appl
15	200.5	11.5	303	2	US-08-294-871A-34	Sequence 34, Appl
16	200.5	11.5	303	2	US-08-294-871A-60	Sequence 60, Appl
17	200.5	11.5	303	3	US-08-876-398A-34	Sequence 34, Appl
18	200.5	11.5	303	3	US-08-876-398A-60	Sequence 60, Appl
19	200	11.5	303	2	US-08-294-871A-8	Sequence 8, Appl
20	200	11.5	303	2	US-08-294-871A-10	Sequence 10, Appl
21	200	11.5	303	2	US-08-294-871A-12	Sequence 12, Appl
22	200	11.5	303	2	US-08-294-871A-22	Sequence 22, Appl
23	200	11.5	303	2	US-08-294-871A-54	Sequence 54, Appl
24	200	11.5	303	3	US-08-876-398A-8	Sequence 8, Appl
25	200	11.5	303	3	US-08-876-398A-10	Sequence 10, Appl
26	200	11.5	303	3	US-08-876-398A-12	Sequence 12, Appl
27	200	11.5	303	3	US-08-876-398A-22	Sequence 22, Appl

#### ALIGNMENTS

RESULT 1  
US-08-294-871A-26  
Sequence 26, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Nanba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294, 871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971, 758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917, 111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211, 641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990

28	200	11.5	303	3	US-08-876-398A-54	Sequence 54, Appl
29	199.5	11.5	303	2	US-08-294-871A-46	Sequence 46, Appl
30	199.5	11.5	303	3	US-08-876-398A-46	Sequence 46, Appl
31	199	11.5	303	1	US-07-917-111-2	Sequence 2, Appl
32	199	11.5	303	1	US-07-917-111-3	Sequence 3, Appl
33	199	11.5	303	1	US-08-479-638-2	Sequence 2, Appl
34	199	11.5	303	1	US-08-479-638-3	Sequence 3, Appl
35	199	11.5	303	2	US-08-294-871A-2	Sequence 2, Appl
36	199	11.5	303	2	US-08-294-871A-14	Sequence 14, Appl
37	199	11.5	303	2	US-08-294-871A-16	Sequence 16, Appl
38	199	11.5	303	2	US-08-294-871A-24	Sequence 24, Appl
39	199	11.5	303	2	US-08-294-871A-28	Sequence 28, Appl
40	199	11.5	303	2	US-08-294-871A-30	Sequence 30, Appl
41	199	11.5	303	2	US-08-294-871A-32	Sequence 32, Appl
42	199	11.5	303	2	US-08-294-871A-52	Sequence 52, Appl
43	199	11.5	303	2	US-08-294-871A-66	Sequence 66, Appl
44	199	11.5	303	3	US-08-876-398A-2	Sequence 2, Appl
45	199	11.5	303	3	US-08-876-398A-14	Sequence 14, Appl

```

1      FILING DATE: 27-DEC-1990
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: JP 078840/1991
4      FILING DATE: 11-APR-1991
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: JP 140051/1991
7      FILING DATE: 12-JUN-1991
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: PCT/JP92/00739
10     FILING DATE: 10-JUN-1992
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: JP 212692/1992
13     FILING DATE: 10-AUG-1992
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: PCT/JP93/01101
16     FILING DATE: 05-AUG-1993
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: JP 340078/1992
19     FILING DATE: 21-DEC-1992
20     ATTORNEY/AGENT INFORMATION:
21     NAME: Wegner, Harold C.
22     REGISTRATION NUMBER: 25, 258
23     REFERENCE/DOCKET NUMBER: 74129/127/AOPB
24     TELECOMMUNICATION INFORMATION:
25     TELEPHONE: (202) 672-5300
26     TELEFAX: (202) 672-5339
27     TELEX: 904136
28     INFORMATION FOR SEQ ID NO: 26:
29     SEQUENCE CHARACTERISTICS:
30     LENGTH: 303 amino acids
31     TYPE: amino acid
32     TOPOLOGY: linear
33     MOLECULE TYPE: protein
34     OS-08-294-871A-26

```

Query Match	11.8%	Score 204	DB 2	Length 303
Best Local Similarity	25.5%	Pred. No. 8.4e-14		
Matches 76	Conservative 46	Mismatches 113	Indels 52	Gaps 12

```

QY 65 EIVOEAAALGCLAFLEP-----AFDFLARNPAEFLILLSEPLNDDLLGGYQSALREC 116
D 29 DMLTKRASKANFLVYFPERLALTTFRRKRNHFDLDEALDSFETE-MRGPVVRPFLEKAAEL 87
QY 117 GIWLSL-----GGEHERGOWEONOKIYNCHVLLNSKGSVVAZYKTHLDVLEIQ 168
D 88 GIGENLGAELVVEGCVKRR-----FNTSILDKSGIKYKRIKTHL-----PGH 1322
QY 159 -----GMR-----EENYRKGGTLEPRVKYTRAGKGLACIDMRPELSLKLAAOGAEILT 220
D 133 KEYEAYRPFQHLERKRYEPBGLGFRVYUDVDAKAKGMPICNDRRMRPEAMRWGLSGAIIIC 1922
QY 221 -YPSAGSVYGPANHEVL-----LRARIESOCVILAAOCSGYNHETRASHGSMV 2711
D 193 GGYNTPTPHNPTVQHDHILTSFNHLLSMQAGSYQNGAMSAAAGKVM-EENOMILGHSCIV 2511
QY 272 DPMGTVAARCE-GRGLSLAIDHFLQOMKONLPV-QNRKRPULG 316
D 252 APTEIVALLTTLEDEVYTAADVDDLRCKREKHEIFENKROHOPNG 298

```

US-08-294-871A-64  
 RESULT 2  
 ; Sequence 64, Application US/08294871A  
 ; Patent No. 5624522  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ikenaka, Yasuhiro  
 ; APPLICANT: Namba, Hirokazu  
 ; APPLICANT: Takano, Masayuki  
 ; APPLICANT: Yamajima, Kazuyoshi  
 ; APPLICANT: Yamada, Yukio  
 ; APPLICANT: Takahashi, Satomi  
 ; APPLICANT: Okubo, Kazuma

```

APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiratah, Yoshio
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentm Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-64

```





APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25, 258  
REFERENCE/DOCKET NUMBER: 74129/130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-876-398A-64

Query Match 11.8%; Score 204; DB 3; Length 303;  
Best Local Similarity 25.7%; Pred. No. 8.4e-14;  
Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

QY 65 ELVOEARLGAFLPE-----AFDFIARNPAETLLSEPLNGDLGGYSQIAREC 116  
DB 29 DMLTFAASRGANFVPELALTFPPRMHFTDEALDSFETE-MGCPVVRPLFEKAEL 87  
QY 117 GIWLSL-----GGFERGODWEDONKIYNCHVLLNSKGSVVASYRTHLCDVEIPQ 168  
DB 88 GIGFNLGYALVEYEGVKRR-----FNTSLIVDKSKGIYVKYRIHL-----PGH 132  
QY 169 -----GPMR--ESNNTKRGGLPEPVKTPACKVGLATCYDMRPFELSLKLAQAGAEILT 220  
DB 133 KKEYARRPQHLKRYFEFGDLGFPYVDVDAKMGAFICNDRRWPEAMRWGLRGAELIT 192  
QY 221 YPSAGSVYGAHMEVL-----LRARIESOCYVIAAOCGRHETTPASYG 266  
DB 193 -----GGYTPFNHPNPIVPGDHLTSPFHLLSMQAGSYNGAAMSAAGKVM--EENMMLG 246  
QY 267 HSMYVDPKCTVVARCSE-GPGICLARIDHFLQOMRQHLPVF-QHRRLPLYG 316  
DB 247 HSCIVAPGELIVALTTLEDEVITAAVDLRCRELEHIFNFKHRORPHYG 298

## RESULT 5

US-08-294-871A-40  
Sequence 40, Application US/08294871A  
Patent No. 5824522

GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Nanba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294, 871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25, 258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-294-871A-40

Query Match 11.7%; Score 203; DB 2; Length 303;  
Best Local Similarity 25.7%; Pred. No. 1.1e-13;  
Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

QY 65 ELVOEARLGAFLPE-----AFDFIARNPAETLLSEPLNGDLGGYSQIAREC 116  
DB 29 DMLTFAASRGANFVPELALTFPPRMHFTDEALDSFETE-MGCPVVRPLFEKAEL 87  
QY 117 GIWLSL-----GGFERGODWEDONKIYNCHVLLNSKGSVVASYRTHLCDVEIPQ 168  
DB 88 GIGFNLGYALVEYEGVKRR-----FNTSLIVDKSKGIYVKYRIHL-----PGH 132  
QY 169 -----GPMR--ESNNTKRGGLPEPVKTPACKVGLATCYDMRPFELSLKLAQAGAEILT 220

Db 133 KEYEAYRPFQHLERKRYEPDGLGFVYVDAAKMGFICNDRRWPEAMRWGLGAEIIC 192  
QY 221 YPSAFGSVTPGAHWEVL-----LRRALIESOCYVIAAOCGRHHETRASYG 266  
Db 193 -----GGYNTPTNPVLPQHDHLTSFHHLLSMOAGSYONGAWSAAKAGM-BEENCMLLG 246  
QY 267 HSNVDPWGTIVARCSE-GRGLCLARIDLHFLQOMROHLPVF-OHRRPDLVYG 316  
Db 247 HSCIVAPTGEIVALTTLLEDEVITAANDLRCRCLREHIFNFKOHPQHYG 298  
RESULT 6  
US-08-876-398A-40  
; Sequence 40, Application US/08876398A  
; Patent No. 6083752  
; GENERAL INFORMATION:  
; APPLICANT: IKENAKA, Yasuhiro  
; APPLICANT: NAKABA, Hirokazu  
; APPLICANT: TAKANO, Masayuki  
; APPLICANT: YAJIMA, Kazuyoshi  
; APPLICANT: YAMADA, Yukio  
; APPLICANT: TAKAHASHI, Satomi  
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/876,398A  
; FILING DATE: 16-JUN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/211,641  
; FILING DATE: 11-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP93/01101  
; FILING DATE: 05-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 340078/1992  
; FILING DATE: 21-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 212692/1992  
; FILING DATE: 10-AUG-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Megner, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 74129/130  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 303 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-876-398A-40  
Query Match 11.7%; Score 203; DB 3; Length 303;  
Best Local Similarity 25.7%; Pred. No. 1,1e-13;  
Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

QY 65 ELVQEARLACGLAFLE-----AFDELARNPAETLLSEPLNDGLGOYSOLAREC 116  
Db 29 DMLTKAASRCRANFLVPELALITFPFRWHTDELDSPETE-MRPPVVRPLFEKAEL 87  
QY 117 GIMLSL-----GCFHERGQDEQONOKIYNCHVILNSKGSVVASRKTHLDVLEIPGQ 168  
Db 88 GIGFNLGVAELVVEGVKRR-----FMSILVDKSGKIKYKRKTHL-----PGH 132  
QY 169 -----GPMR--ESNYTRPGSTLEPPVKTGAKVGLAICIDMRPELSLKAQAGAILT 220  
Db 133 KEYEAYRPFQHLERKRYEPDGLGFVYVDAAKMGFICNDRRWPEAMRWGLGAEIIC 192  
QY 221 YPSAFGSVTPGAHWEVL-----LRRALIESOCYVIAAOCGRHHETRASYG 266  
Db 193 -----GGYNTPTNPVLPQHDHLTSFHHLLSMOAGSYONGAWSAAKAGM-BEENCMLLG 246  
QY 267 HSNVDPWGTIVARCSE-GRGLCLARIDLHFLQOMROHLPVF-OHRRPDLVYG 316  
Db 247 HSCIVAPTGEIVALTTLLEDEVITAANDLRCRCLREHIFNFKOHPQHYG 298  
RESULT 7  
US-08-294-871A-6  
; Sequence 6, Application US/08294871A  
; Patent No. 5824522  
; GENERAL INFORMATION:  
; APPLICANT: IKENAKA, Yasuhiro  
; APPLICANT: NAKABA, Hirokazu  
; APPLICANT: TAKANO, Masayuki  
; APPLICANT: YAJIMA, Kazuyoshi  
; APPLICANT: YAMADA, Yukio  
; APPLICANT: TAKAHASHI, Satomi  
; APPLICANT: OKUBO, Kazuma  
; APPLICANT: YAMADA, Kazuhiko  
; APPLICANT: HIRASHI, Yoshiro  
; TITLE OF INVENTION: Immobilized Enzyme Preparation and  
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/294,871A  
; FILING DATE: 22-AUG-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/971,758  
; FILING DATE: 12-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/917,111  
; FILING DATE: 07-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/211,641  
; FILING DATE: 11-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 400848/1990  
; FILING DATE: 07-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP91/01696  
; FILING DATE: 06-DEC-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 407922/1990  
; FILING DATE: 27-DEC-1990  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-294-871A-6

Query Match 11.6%; Score 202; DB 2; Length 303;  
Best Local Similarity 25.7%; Pred. No. 1.4e-13;

Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

QY 65 ELVOEAAKRLGACLAFLPE-----AFDFIARNPAETLLSEPLNGDLLGOYSQALREC 116  
DB 29 DMLTKAASRGANFIVPELALTTFFPRMHTDEALDSFYETE-MGCPVVRPLFEKAAEL 87  
QY 117 GVLWLSL-----GGFHERGQDMEONOKIYNCHVLLNSKGSVYASVYRKTLCDELVEIPQ 168  
DB 88 GIGFNLGYAELVEGCVKRR-----FNSTILVDSKGIYGYRKIHL-----PGH 132  
QY 169 -----GPMR--ESNTKRGCTLEPPVKTGACVGLAICVDMFPELSKLAAAGAEIIT 220  
DB 133 KEYEAYRPFQHLERKRFEPEDLGFVYVDAAKMGFICNDRMPEAMVYMGIRGAELIC 192  
QY 221 YPSAFGSVTPGAHWEVL-----LRAAIESQCYVIAAOCGRHHETRASVG 266  
DB 193 -----GGYMTPTNPLVPQHDLTSTFHHLLSMQAGSYQNGAMSAAGKVGKGM-EECNMLLG 246  
QY 267 HSMNVDPKGTIVARCS-GRGLCLARIDLHFLQOMQHLPVF-QHRRPDLYG 316  
DB 247 HSCIVAPTEIGVIALTTTLEDEVITTAVDLDRCRELREHIFNFQKHOPQHYG 298

RESULT 8

US-08-876-398A-6  
Sequence 6, Application US/08876398A  
Patent No. 6083752

GENERAL INFORMATION:

APPLICANT: IKENAKA, Yasuhiro  
APPLICANT: NAMBA, Hltokazu  
APPLICANT: TAKANO, Masayoshi  
APPLICANT: YAJIMA, Kazuyoshi  
APPLICANT: YAMADA, Yukio  
APPLICANT: TAKAHASHI, Satomi  
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF  
NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876,398A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-876-398A-6

Query Match 11.6%; Score 202; DB 3; Length 303;  
Best Local Similarity 25.7%; Pred. No. 1.4e-13;

Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

QY 65 ELVOEAAKRLGACLAFLPE-----AFDFIARNPAETLLSEPLNGDLLGOYSQALREC 116  
DB 29 DMLTKAASRGANFIVPELALTTFFPRMHTDEALDSFYETE-MGCPVVRPLFEKAAEL 87  
QY 117 GVLWLSL-----GGFHERGQDMEONOKIYNCHVLLNSKGSVYASVYRKTLCDELVEIPQ 168  
DB 88 GIGFNLGYAELVEGCVKRR-----FNSTILVDSKGIYGYRKIHL-----PGH 132  
QY 169 -----GPMR--ESNTKRGCTLEPPVKTGACVGLAICVDMFPELSKLAAAGAEIIT 220  
DB 133 KEYEAYRPFQHLERKRFEPEDLGFVYVDAAKMGFICNDRMPEAMVYMGIRGAELIC 192  
QY 221 YPSAFGSVTPGAHWEVL-----LRAAIESQCYVIAAOCGRHHETRASVG 266  
DB 193 -----GGYMTPTNPLVPQHDLTSTFHHLLSMQAGSYQNGAMSAAGKVGKGM-EECNMLLG 246  
QY 267 HSMNVDPKGTIVARCS-GRGLCLARIDLHFLQOMQHLPVF-QHRRPDLYG 316  
DB 247 HSCIVAPTEIGVIALTTTLEDEVITTAVDLDRCRELREHIFNFQKHOPQHYG 298

RESULT 9

US-08-294-871A-44  
Sequence 44, Application US/08294871A  
Patent No. 5824522

```

GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuniko
APPLICANT: Hiraishi, Yoshiko
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136

```

```

; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-294-871A-44

Query Match 11.6%; Score 201.5; DB 2; Length 303;
Best Local Similarity 25.8%; Pred. No. 1,6e-13;
Matches 77; Conservative 44; Mismatches 115; Indels 63; Gaps 13;

QY 50 VTSPNKQENFKTCALVEQPARAGACLAIPFAFDIARPAFTLLSEPLNDLLGQY 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 31 LTKASRGANFIYVPEL-----ALTFFPRMY-FTDEALDSFYETE-MPGVVRPL 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 SOLARECGIMVLSL-----GGFHERGQDWEQNOCKLYNCVLLNSKGSVYASRYKTHLC 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 81 FERAELGIGFNLGYALVYEGYKRR-----FNTSLVDKSGKIVGKRYKTHL- 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 DVEIPGQ-----GPMR-ESNYTKPGTLEPPVKTPEGKVGALICYDMRPPELSLKLQ 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 ---PGHKEYEAYRPFQHLERKRYEPDGLGPPYVDVDAKKGMFTICNDRRMPEAMRVWGL 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 214 AGAEILTPPSAFSGVTPARHEVL-----LRAALESCCYIAAQCGRHH 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 RGAELIC-----GGYNPTNPPLVPOHDHLTSFHHLLSMQAGSYQNGAWSAAAGKGM-E 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 ETRASGSHMVDPMGTGVVARCSE-GPGLCLARIDLHFLQOMROHLPVF-OHRRPDLYG 316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 ENCMILGSHCVAPGTGELVALLTLEDEVITAAVDLDRCRRLRHHNFKHROPQHYG 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-08-876-398A-44
; Sequence 44, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAMBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992

```





CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876.398A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211.641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-876-398A-42

Query Match 11.6%, Score 201, DB 3; Length 303;  
Best Local Similarity 26.1%; Pred. No.1.8e-13;  
Matches 75; Conservative 47; Mismatches 113; Indels 52; Gaps 12;

QY 65 ELVQDAARLGCALFLPE-----AFDFIARNPAETILLSEPLNGDLLGQYSOLAREC 116  
DB 29 DMLTKAASRGANFIVPELPALTFPPRWMHFTDEALDSFYETE-MGPVVRPLFEKAEL 87  
QY 117 GIWLSL-----GGFHERGQDWEQONKIYNCHVLNSKGSVASYRTHLCDVEIPQ 168  
DB 88 GIGFNLGALVELVEGVGKRR-----FNTSLVDKSGKIVKRYKRIHL-----PGH 132  
QY 169 -----GPMR--ESNYTKRPGTLEPPVKTGKVGLAICYDMRPELSIKLAQAQAEIIT 220  
DB 133 KEYEAYRPFQHLERKRYFEPGDLGFVYVDADAKMGMFICNDRMPDEAWRMVGLRGAEIITC 192  
QY 221 --YPSAFSGVTGRAHWEVL-----LRARATESOCYVIAAOCGRNHETRASYSYSHMVY 271  
DB 153 GGYNTPHNPSVPQHDHLTSFHHLSMQAGSYONGAMSAAGKGM-EENCMLLGHSCIV 251  
QY 272 DPMGVNVARCSE-GPGICLARIDLHFLQOMRQHLPVF-QHRRPDLYG 316  
DB 252 APTGELVALTTTLEDEVITAAVDLDRCRELREHIFNFKQROPHNG 298

RESULT 14  
US-08-876-398A-62  
Sequence 62, Application US/08876398A  
Patent No. 6083752  
GENERAL INFORMATION:  
APPLICANT: IKENAKA, Yasuhiro  
APPLICANT: NANBA, Hirokazu  
APPLICANT: TAKANO, Masayuki  
APPLICANT: YAJIMA, Kazuyoshi  
APPLICANT: YAMADA, Yukio  
APPLICANT: TAKAHASHI, Satomi  
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
TITLE OF INVENTION: THERMOSTABILITY AND USE THEROF  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER

STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876.398A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211.641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-876-398A-62

Query Match 11.6%, Score 201, DB 3; Length 303;  
Best Local Similarity 25.7%; Pred. No.1.8e-13;  
Matches 75; Conservative 44; Mismatches 111; Indels 62; Gaps 12;

QY 65 ELVQDAARLGCALFLPE-----AFDFIARNPAETILLSEPLNGDLLGQYSOLAREC 116  
DB 29 DMLTKAASRGANFIVPELPALTFPPRWMHFTDEALDSFYETE-MGPVVRPLFEKAEL 87  
QY 117 GIWLSL-----GGFHERGQDWEQONKIYNCHVLNSKGSVASYRTHLCDVEIPQ 168  
DB 88 GIGFNLGALVELVEGVGKRR-----FNTSLVDKSGKIVKRYKRIHL-----PGH 132  
QY 169 -----GPMR--ESNYTKRPGTLEPPVKTGKVGLAICYDMRPELSIKLAQAQAEIIT 220  
DB 133 KEYEAYRPFQHLERKRYFEPGDLGFVYVDADAKMGMFICNDRMPDEAWRMVGLRGAEIITC 192  
QY 221 YPSAFSGVTGRAHWEVL-----LRARATESOCYVIAAOCGRNHETRASYSY 266  
DB 193 -----GGYNTPHNPSVPQHDHLTSFHHLSMQAGSYONGAMSAAGVGM-EENCMLLG 246  
QY 267 HSMVVDPMGVNVARCSE-GPGICLARIDLHFLQOMRQHLPVF-QHRRPDLYG 316  
DB 247 HSCIVAPTEGELVALTTTLEDEVITAAVDLDRCRELREHIFNFKQROPHNG 298

RESULT 15  
US-08-294-871A-34  
Sequence 34, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro



```

; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-294-871A-34

```

Query Match	11.5%;	Score 200.5;	DB 2;	Length 303;
Best Local Similarity	25.8%;	Pred. No. 2e-13;		
Matches 77; Conservative	44;	Mismatches 115;	Indels 63;	Gaps 13;

QY	50	VTSPNNQENKTCAEIVOEAAARJGACIAIEPFEFDEIARNPATLLISEPNDLLGOY	109
Db	31	LTKAASGANIYPERL-----ALTFFPRMY-ETDEAELDSFYETE-MPGYVVRPL	80
QY	110	SQARREGCIMISSL-----GCFHEGQDMEQNOKIYNCHVLLNSGVSVAATRHLIC	166
Db	81	FEKAAEIGIGENGLYAEIVVEGSKRR-----FNTSLDKSGKIGXKRYKTHL-	123
QY	162	DVEIRPGQ-----GPMR--ESNYTKPGGLEPVPYKTPAGKVGALICYDMRPEPSTLKAQ	213
Db	130	-----PGHKEYATYRPFQHLERKRYTERBDLGFPPYVDVDAKKMGMTICDRKRPRFEMRYMGL	188
QY	214	AGAEILTYPSAFSGVTGRAHWEVL-----LRARATESOCUYTLAAQSGRNH	255
Db	186	RGAEIIC-----GGYNTPTNHPILVPOHDHILTSFNHLLSMQSGSYQAMSAALAAKVGUM-E	239
QY	260	ETRASVSGHSMVVDWGVYVAAQCSB-GRGLCLARLDLHFLQMRONHLVVF--ONHRRPDLXG	316
Db	240	ENCMILGKISCVATPGLVETLITLLEDEVYITAAVDLDRCEFLRHINFEKQHNORPONG	298

Search completed: April 26, 2002, 16:55:12  
Job time: 255 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:55:12 ; Search time 28.88 Seconds  
(without alignments)  
282.070 Million cell updates/sec

Title: US-09-357-675C-25

Perfect score: 1921  
Sequence: 1 PLAAXLADRPDPRRLRMV.....LPVGHRRPDLYGNLGHPLS 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 212252 segs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCrUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.5	10.4	303	2	US-08-294-871A-64
2	200.5	10.4	303	3	US-08-876-398A-64
3	199.5	10.4	303	2	US-08-294-871A-26
4	199.5	10.4	303	2	US-08-294-871A-40
5	199.5	10.4	303	2	US-08-294-871A-44
6	199.5	10.4	303	2	US-08-876-398A-26
7	199.5	10.4	303	3	US-08-876-398A-40
8	199.5	10.4	303	3	US-08-876-398A-44
9	198.5	10.3	303	2	US-08-294-871A-6
10	198.5	10.3	303	2	US-08-294-871A-34
11	198.5	10.3	303	2	US-08-294-871A-60
12	198.5	10.3	303	3	US-08-876-398A-6
13	198.5	10.3	303	3	US-08-876-398A-34
14	198.5	10.3	303	3	US-08-876-398A-60
15	197.5	10.3	303	2	US-08-294-871A-62
16	197.5	10.3	303	3	US-08-876-398A-62
17	196.5	10.2	303	2	US-08-294-871A-10
18	196.5	10.2	303	2	US-08-294-871A-12
19	196.5	10.2	303	2	US-08-294-871A-22
20	196.5	10.2	303	2	US-08-294-871A-38
21	196.5	10.2	303	2	US-08-294-871A-42
22	196.5	10.2	303	2	US-08-294-871A-46
23	196.5	10.2	303	2	US-08-294-871A-54
24	196.5	10.2	303	2	US-08-294-871A-58
25	196.5	10.2	303	2	US-08-294-871A-68
26	196.5	10.2	303	3	US-08-876-398A-10
27	196.5	10.2	303	3	US-08-876-398A-12

28	196.5	10.2	303	3	US-08-876-398A-22	Sequence 22, Appl
29	196.5	10.2	303	3	US-08-876-398A-38	Sequence 38, Appl
30	196.5	10.2	303	3	US-08-876-398A-42	Sequence 42, Appl
31	196.5	10.2	303	3	US-08-876-398A-46	Sequence 46, Appl
32	196.5	10.2	303	3	US-08-876-398A-54	Sequence 54, Appl
33	196.5	10.2	303	3	US-08-876-398A-58	Sequence 58, Appl
34	196.5	10.2	303	3	US-08-876-398A-68	Sequence 68, Appl
35	195.5	10.2	303	1	US-07-917-111-2	Sequence 2, Appl
36	195.5	10.2	303	1	US-07-917-111-3	Sequence 3, Appl
37	195.5	10.2	303	1	US-08-479-638-2	Sequence 2, Appl
38	195.5	10.2	303	1	US-08-479-638-3	Sequence 3, Appl
39	195.5	10.2	303	2	US-08-294-871A-2	Sequence 2, Appl
40	195.5	10.2	303	2	US-08-294-871A-4	Sequence 4, Appl
41	195.5	10.2	303	2	US-08-294-871A-8	Sequence 8, Appl
42	195.5	10.2	303	2	US-08-294-871A-14	Sequence 14, Appl
43	195.5	10.2	303	2	US-08-294-871A-16	Sequence 16, Appl
44	195.5	10.2	303	2	US-08-294-871A-18	Sequence 18, Appl
45	195.5	10.2	303	2	US-08-294-871A-20	Sequence 20, Appl

## ALIGNMENTS

RESULT 1  
US-08-294-871A-64  
Sequence 64, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Nanba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
PROCESS for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-APR-1994  
CLASSIFICATION: A35  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/921,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990



OY 344 PVF-OHRRDLXG 355  
DB 286 FNFKOHPOHYG 298

## RESULT 3

US-08-294-871A-26  
Sequence 26, Application US/08294871A

Patent No. 5824522

GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro

APPLICANT: Namba, Hirokazu

APPLICANT: Takano, Masayuki

APPLICANT: Yajima, Kazuyoshi

APPLICANT: Yamada, Yukio

APPLICANT: Takahashi, Satomi

APPLICANT: Okubo, Kazuma

APPLICANT: Yamada, Kazuhiko

APPLICANT: Hiraiishi, Yoshiro

TITLE OF INVENTION: Immobilized Enzyme Preparation and  
Process for Producing D-a-Amino Acid

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/294,871A

FILING DATE: 22-AUG-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/971,758

FILING DATE: 12-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917,111

FILING DATE: 07-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/211,641

FILING DATE: 11-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 400848/1990

FILING DATE: 07-DEC-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP91/01696

FILING DATE: 06-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 407922/1990

FILING DATE: 27-DEC-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 078840/1991

FILING DATE: 11-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 140051/1991

FILING DATE: 12-JUN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP92/00739

FILING DATE: 10-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 212692/1992

FILING DATE: 10-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/01101

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

COMPUTER READABLE FORM:

APPLICATION NUMBER: JP 340078/1992

FILING DATE: 21-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wegner, Harold C.

REGISTRATION NUMBER: 25,258

REFERENCE/DOCKET NUMBER: 74129/127/AOPA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 303 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-294-871A-26

Query Match

Best Local Similarity 26.98; Pred. No. 2.1e-13;

Matches 83; Conservative 47; Mismatches 129; Indels 49; Gaps 14;

OY 83 LVAVCQ---VTSPDRQNEFKTCAELVREARLGLAFLPE--AFDFIAR---DPAET 133

DB 5 ILAVGOGGPIARRETRQVVRLLDMLTKRASGAMPVPELALTTFFRRMFTDEAEL 64

OY 134 LHLSE-PLGKLLLEETQLARECGMLSLGFFH---ERGQDWEOTOKIYNCHVILNSKA 189

DB 65 DSFEYTEMFGPVYRPLFEKRAELIGFNLGYALVVEGG-----VKRRFTSLIVDSK 119

OY 190 VVATYRKTLCVETPGQGMCSNSTMPGSLSEYSPF-----AKIGLAVC 238

DB 120 IYKRYKRIHL-----PGHK---EYEAIVRPQHLKRYFEFGDLGFPYVDVDAKMGAFIC 171

OY 239 YDRPPELSLALAQAGAEILF--YPSAFSGITSPAHMEVL-----LRARLETQCYV 289

DB 172 NDRMPPEANRWMLRGAEITCGSYNTPTNPTVPQDHLTSPHLLSMQSGYQNGAMSA 231

OY 290 AAAQCGRHHEKRASYSYGNVDPWGTIVARCSE--GGELCLARIDLNTQLRRHLLPVF-Q 347

DB 232 AAGKVGCM--EENCMLLGHSCIVAPTEIVALTTLLEDEVITAAVDDLDRCLRHIFNKQ 290

OY 348 HRRPDLXG 355

DB 291 HROPPOHYG 298

RESULT 4

US-08-294-871A-40

Sequence 40, Application US/08294871A

Patent No. 5824522

GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro

APPLICANT: Namba, Hirokazu

APPLICANT: Takano, Masayuki

APPLICANT: Yajima, Kazuyoshi

APPLICANT: Yamada, Yukio

APPLICANT: Takahashi, Satomi

APPLICANT: Okubo, Kazuma

APPLICANT: Yamada, Kazuhiko

APPLICANT: Hiraiishi, Yoshiro

TITLE OF INVENTION: Immobilized Enzyme Preparation and

Process for Producing D-a-Amino Acid

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.

ZIP: 20007-5109

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```

```

Query Match      10.4%: Score 199.5: DB 2: Length 303:
Best Local Similarity 26.2%: Pred. No. 2.1e-13:
Matches 82: Conservative 46: Mismatches 126: Indels 59: Gaps 14:

QY 83 LVAVQ---VSTPDKQNFKTCALVEAEARLGAFLPE--AFDEIAR----DPAET 133
      ::|||::: ||| ||| ||| |||
Db 5 ILAVGQGPPIARAETREQVVRLLDMLPKRASNFIPEELALTFPPRHNFDEAL 64
      ::|||::: ||| ||| ||| |||

QY 134 LHLSE-PLGKLLLEETQLAREGGLMSLGGH---ERQDMEQJOKIYNCHVLLSKSA 189
      ::|||::: ||| ||| ||| |||
Db 65 DSYETETMPGVVRLPEKRAELGIGFNLGYALVEVG---VKRRNTSILVDKSK 119
      ::|||::: ||| ||| ||| |||

```

[illegible]

```

1      RESULT 5
2      US-08-294-871A-44
3      : Sequence 44, Application US/08294871A
4      : Patent No. 5824522
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Ikenaka, Yasuhiro
9      : APPLICANT: Nanba, Hirokazu
10     : APPLICANT: Takano, Masayuki
11     : APPLICANT: Yajima, Kazuyoshi
12     : APPLICANT: Yamada, Yukio
13     : APPLICANT: Takahashi, Satomi
14     : APPLICANT: Okubo, Kazuma
15     : APPLICANT: Hiraishi, Yoshiro
16     : TITLE OF INVENTION: Immobilized Enzyme Preparation and
17     : TITLE OF INVENTION: Process for Producing D-a-Amino Acid
18     : NUMBER OF SEQUENCES: 70
19     :
20     : CORRESPONDENCE ADDRESS:
21     : ADDRESSEE: Foley & Lardner
22     : STREET: 3000 K Street, N.W.
23     : City: Washington
24     : STATE: D.C.
25     : COUNTRY: U.S.
26     : ZIP: 20007-5109
27     :
28     : COMPUTER READABLE FORM:
29     :
30     : MEDIUM TYPE: Floppy disk
31     : COMPUTER: IBM PC compatible
32     : OPERATING SYSTEM: PC-DOS/MS-DOS
33     : SOFTWARE: PatentIn Release #1.0, Version #1.25
34     : CURRENT APPLICATION DATA:
35     : APPLICATION NUMBER: US/08/294,871A
36     : FILING DATE: 22-APR-1994
37     : CLASSIFICATION: 435
38     :
39     : PRIOR APPLICATION DATA:
40     : APPLICATION NUMBER: US 07/971,758
41     : FILING DATE: 12-APR-1993
42     : PRIOR APPLICATION DATA:
43     : APPLICATION NUMBER: US 07/917,111
44     : FILING DATE: 07-AUG-1992
45     : PRIOR APPLICATION DATA:
46     : APPLICATION NUMBER: US 08/211,641
47     : FILING DATE: 11-APR-1994
48     : PRIOR APPLICATION DATA:
49     : APPLICATION NUMBER: JP 400848/1990
50     : FILING DATE: 07-DEC-1990
51     : PRIOR APPLICATION DATA:
52     : APPLICATION NUMBER: PCT/JP91/01696
53     : FILING DATE: 06-DEC-1991
54     : PRIOR APPLICATION DATA:
55     : APPLICATION NUMBER: JP 407922/1990
56     : FILING DATE: 27-DEC-1990
57     : PRIOR APPLICATION DATA:
58     : APPLICATION NUMBER: JP 078840/1991
59     : FILING DATE: 11-APR-1991
60     : PRIOR APPLICATION DATA:
61     : APPLICATION NUMBER: JP 140051/1991

```









Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Namba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiko  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399

TELEX: 904136  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-294-871A-34  
Query Match 10.3%; Score 198.5; DB 2; Length 303;  
Best Local Similarity 26.2%; Pred. No. 2,7e-13;  
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;  
QY 83 LVAVCO---VTSPPDKQONFKTCALVREARGLACLAFLPE--APDFIAR---DPAET 133  
DB 5 ILAVGQGPRIARETRERQVAVRLLDMLTAAASGANFIYFPELALTTFFPRWFTDEABL 64  
QY 134 LHLSE-PLGGKLELEYTQLARECGMLSLGFEH---ERGDMEQITOKIYNCHVLLNSKA 189  
DB 65 DSFEYTEMFGPVYRPLFEKAAELGIFNLGYAELVVEGG---VKRRNTSILVDKSGK 119  
QY 190 VVATYRKTHLDVEITPGQPMCESNSTMPGSLSEVSTP-----ACKIGLAVC 238  
DB 120 IVGKYRKRIHL---PGHK---EYEAIRPQHLKRYFEPGDLGFPYVDVDAKMGMFIC 171  
QY 239 YDMRPEPLSLALQAAGAEILTPSAGFSITGPAHMEVL-----LRARAIET 284  
DB 172 NDRMPEANRVMGLRGAEIITC-----GYNTPTPHNPLVQHDHLTFSFHLLSQAQSYON 226  
QY 285 OCYVVAAGCGRRHERRASYSMVVDPMGTVAARCSE-GPGICLARIDLNTYLRQLRRHL 343  
DB 227 GAWSAAGKVMG-EENCMLGHSCIVAPGGEIYALFTTLEDEVITFAAVDIDRCREIREHI 285  
QY 344 PVF-QHRPDLTG 355  
DB 286 FNEKHROPQHYG 298  
RESULT 11  
US-08-294-871A-60  
Sequence 60, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Namba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiko  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/130/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-294-871A-60

Query Match 10.3%; Score 198.5; DB 2; Length 303;  
Best Local Similarity 26.2%; Pred. No. 2,7e-13;  
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 83 LVAVCQ---VSTPDKQONFTCAELVREARLACLAFLPE--AFDFIAR---DPAET 133  
DB 5 ILAVGQGFPIARAETREQVAVRLLDMLTKASRGANFIVPELALTTFFPRMHTDEAEL 64  
QY 134 LHUSE-PLGGLILEEYQTLAREGLMLSLGFGH--ERQDWEQOTOKLYNCHVLNLSKA 189  
DB 65 DSFEYTEMPPGVVRPLLEKKAELGIGRLGALVELVEGG-----VKRFNLSILVDSKG 119  
QY 190 VVATYRTHLCDVEIPGQGPWCESNTPGSPLESVPSTP-----AGKIGLAVC 238  
DB 120 IVGKYRKRIHT-----PGHK---EYEAAYRPFQHLERKRYFEPDGLGFPVYDVDAKMGFIC 171  
QY 239 YDMFPFELSLALAGAGAEILITPFAFGSITPRAHVEYL-----LARALET 284  
DB 172 NDRMWPRAWRMYGLRGAEIITC-----GGYNTPTNHPPLVPQHDHLTSPHLLSMDAGSYON 226  
QY 285 QCYVVAAGCGRHHEKASYSYGSMSVVDPMGTIVVARCSE-GRGLCLARIDLNYLQLRRLH 343

DB 227 GAWSAAGKSGM-BENCMLLGHSCIVAPTGEIVALTTLTLEDEVITAAVDLDRCELREHI 285  
QY 344 PVF-QHRPDLXG 355  
DB 286 FNFQKHROPQHYG 298

RESULT 12  
US-08-876-398A-6  
Sequence 6, Application US/08876398A  
Patent No. 6083752  
GENERAL INFORMATION:  
APPLICANT: IKENAKA, Yasuhiro  
APPLICANT: NAKABA, Hirokazu  
APPLICANT: TAKANO, Masayuki  
APPLICANT: YAJIMA, Kazuyoshi  
APPLICANT: YAMADA, Yukio  
APPLICANT: TAKAHASHI, Satomi  
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
City: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876,398A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-876-398A-6

Query Match 10.3%; Score 198.5; DB 3; Length 303;  
Best Local Similarity 26.2%; Pred. No. 2,7e-13;  
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 83 LVAVCQ---VSTPDKQONFTCAELVREARLACLAFLPE--AFDFIAR---DPAET 133  
DB 5 ILAVGQGFPIARAETREQVAVRLLDMLTKASRGANFIVPELALTTFFPRMHTDEAEL 64

QY 134 LHLSE-PLGGKLEEYTOAREGGLWSLGFH---ERGQDWEQTOIKYINCHVLSKGA 189  
DB 65 DSEYETEMGPVVRPLFEKAAELGIGFNGLYAEVYEGG-----VKRRNTSLVDKSKR 119  
QY 190 VVATYRKTHLCDVEIPGOGPCESNSTMPGSLSPVSTP-----AGKIGLAVC 238  
DB 120 IVGKYRKHL-----PGHK---EYEAIRPFOHLEKRYFEPGDLGFPVYDVDAKMGKMFIC 171  
QY 239 YMRPEPESLALAOAGAEILTYPSAFGSITGPAHMEVL-----LRARAIFET 284  
DB 172 NDRRPEARVWGLRGAEITIC-----GGYNTPTNHPVQHDHLSFHHLSMQAGSYON 226  
QY 285 QCYVVAACGRRHHERKASYSYGHSMVDPGTVVARCSE-GPGICLARIDLNTYLRQLRRHL 343  
DB 227 GAMSAAAGKVGW-EENCMLLGHSCIVAPTEIYALTTLEDEYITTAAYDLCRCRELREHI 285  
QY 344 PVF-QHRRPDLG 355  
DB 286 FNFKQHRQPHYG 298

RESULT 13  
US-08-876-398A-34  
; Sequence 34, Application US/08876398A  
; Patent No. 6083752  
; GENERAL INFORMATION:  
; APPLICANT: IKENAKA, Yasuhiro  
; APPLICANT: NANBA, Hirokazu  
; APPLICANT: TAKANO, Masayuki  
; APPLICANT: YAJIMA, Kazuyoshi  
; APPLICANT: YAMADA, Yukio  
; APPLICANT: TAKAHASHI, Satomi  
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/876,398A  
; FILING DATE: 16-JUN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/211,641  
; FILING DATE: 11-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP93/01101  
; FILING DATE: 05-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 340078/1992  
; FILING DATE: 21-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 212692/1992  
; FILING DATE: 10-AUG-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wegner, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 74129/130  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
;

LENGTH: 303 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-876-398A-34

Query Match 10.3%; Score 198.5; DB 3; Length 303;  
Best Local Similarity 26.2%; Pred. No. 2.7e-13;  
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 83 LVAVCO---VTSTPRKQKQFKCAELVREARLGAFLPE--AFDFIAR---DPAET 133  
DB 5 ILAVGQGPRIARAEFREQVAVRLMLTKAASGANFIVEPELATTFPPRRYFTDEAEI 64  
QY 134 LHLSE-PLGGKLEEYTOAREGGLWSLGFH---ERGQDWEQTOIKYINCHVLSKGA 189  
DB 65 DSEYETEMGPVVRPLFEKAAELGIGFNGLYAEVYEGG-----VKRRNTSLVDKSKR 119  
QY 190 VVATYRKTHLCDVEIPGOGPCESNSTMPGSLSPVSTP-----AGKIGLAVC 238  
DB 120 IVGKYRKHL-----PGHK---EYEAIRPFOHLEKRYFEPGDLGFPVYDVDAKMGKMFIC 171  
QY 239 YMRPEPESLALAOAGAEILTYPSAFGSITGPAHMEVL-----LRARAIFET 284  
DB 172 NDRRPEARVWGLRGAEITIC-----GGYNTPTNHPVQHDHLSFHHLSMQAGSYON 226  
QY 285 QCYVVAACGRRHHERKASYSYGHSMVDPGTVVARCSE-GPGICLARIDLNTYLRQLRRHL 343  
DB 227 GAMSAAAGKVGW-EENCMLLGHSCIVAPTEIYALTTLEDEYITTAAYDLCRCRELREHI 285  
QY 344 PVF-QHRRPDLG 355  
DB 286 FNFKQHRQPHYG 298

RESULT 14  
US-08-876-398A-60  
; Sequence 60, Application US/08876398A  
; Patent No. 6083752  
; GENERAL INFORMATION:  
; APPLICANT: IKENAKA, Yasuhiro  
; APPLICANT: NANBA, Hirokazu  
; APPLICANT: TAKANO, Masayuki  
; APPLICANT: YAJIMA, Kazuyoshi  
; APPLICANT: YAMADA, Yukio  
; APPLICANT: TAKAHASHI, Satomi  
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/876,398A  
; FILING DATE: 16-JUN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/211,641  
; FILING DATE: 11-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP93/01101  
; FILING DATE: 05-AUG-1993  
; PRIOR APPLICATION DATA:  
;

```

1  APPLICATION NUMBER: JP 340078/1992
2  FILING DATE: 21-DEC-1992
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER: JP 212692/1992
5  FILING DATE: 10-AUG-1992
6  ATTORNEY/AGENT INFORMATION:
7  NAME: Wegner, Harold C.
8  REGISTRATION NUMBER: 25,258
9  REFERENCE/DOCKET NUMBER: 74129/130
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: (202) 672-5300
12 TELEFAX: (202) 672-5399
13 INFORMATION FOR SEQ ID NO: 60:
14 SEQUENCE CHARACTERISTICS:
15     LENGTH: 303 amino acids
16     TYPE: amino acid
17     TOPOLOGY: linear
18 MOLECULE TYPE: protein
19
20 US-08-876-398A-60

```

```

Db      65  DSEFTEEMPGPVVRPLFEKAELIGFNLGYAELVYEGC-----VKRENTSILVDKSGK 119
QY      190  VVATYRKTHLCDVEIPGOGMCESNSTMPGSPLESVPSTP-----AGKIGLAVC 238
      :!|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      120  IVGKYRKIH-----PGHK---EYEA YRPQHLEKRYFERGDLGFVPYDADAKMGMFIC 171
QY      239  YDMRPELSIALAQAGAEILTYPSAFSGITGPAHWEVL-----LRARALET 284
      :!|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      172  NDRRMEPEAMRVVMGLRGAELIC-----GGYNTPTNPNPVPQHDHLLTFHHLLSMQAGSYON 226
QY      285  QCVVVAACGGRHHEKRASYGSHMVDPMPGTVVARCSE-GPGCLARIDILNYLRLRHL 343
Db      227  GAWSAAGKVGW-EENCMLGHSCIVAPTEIYALTTLEDEVITAAVDLDRCRELREHI 285
QY      344  PVF-QHRRPDLYG 355
      :|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      286  FNFQKHROPQHYG 298

```

Search completed: April 26, 2002, 16:55:12  
 Job time: 255 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 17:02:23 ; Search time 57.61 Seconds  
(without alignments)  
919.121 Million cell updates/sec

Title: US-09-357-675C-25

Perfect score: 1921

Sequence: 1 PLAAAXLAPDRPRDRLRMV.....LPVFQHRRLDLYGNLGHPLS 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp\_invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1743	90.7	327	4	076091	076091 homo sapien
2	1482	77.1	323	11	088526	088526 mus musculu
3	1479	77.0	323	11	09R1M4	09R1M4 mus musculu
4	1372	71.4	290	11	Q9DBF2	Q9DBF2 mus musculu
5	956	49.8	288	13	Q918G0	Q918G0 xenopus lae
6	710.5	37.0	460	5	076464	076464 drosophila
7	702	36.5	440	5	076463	076463 caenorhabd1
8	645.5	33.6	316	10	Q91E50	Q91E50 arabidopsis
9	575.5	30.0	276	3	094660	094660 schizosacch
10	537.5	28.0	272	1	027839	027839 methanobact
11	497	25.9	282	2	Q9HVU6	Q9HVU6 pseudomonas
12	485.5	25.3	276	4	Q9NOR4	Q9NOR4 homo sapien
13	479	24.9	270	2	083040	083040 plectonema
14	474.5	24.7	276	11	Q9JHW2	Q9JHW2 mus musculu
15	436.5	22.7	275	2	Q9K0U4	Q9K0U4 vibrio chol
16	433.5	22.6	283	2	Q9A9Y3	Q9A9Y3 caulobacter
17	406	21.1	318	10	Q9LYH1	Q9LYH1 arabidopsis
18	401	20.9	270	2	Q9J7I3	Q9J7I3 neisseria m
19	398	20.7	270	2	Q9K0W5	Q9K0W5 neisseria m

20	397	20.7	113	6	029367	029367 sus scrofa
21	395.5	20.6	189	11	09CTG9	09CTG9 mus musculu
22	387	20.1	283	5	09VHE4	09VHE4 drosophila
23	371	19.3	231	11	09D0S1	09D0S1 mus musculu
24	342.5	17.8	264	2	Q9XA70	Q9XA70 streptomyce
25	318	16.6	297	2	Q9RRF7	Q9RRF7 deinococcus
26	317	16.5	270	1	09H1W8	09H1W8 thermoplas
27	312	16.2	257	1	030121	030121 archaeoglob
28	291.5	15.2	300	10	09XG19	09XG19 lycopersico
29	279	14.5	287	1	09Y9L1	09Y9L1 aeropyrum p
30	274.5	14.3	271	2	09KE11	09KE11 bacillus ha
31	273.5	14.2	292	2	0916J8	0916J8 pseudomonas
32	264.5	13.8	295	2	09PA02	09PA02 xylella fas
33	262	13.6	270	2	091241	091241 pseudomonas
34	255.5	13.3	259	2	031664	031664 bacillus su
35	253.5	13.2	282	2	09Z505	09Z505 zymomonas m
36	252	13.1	298	12	089413	089413 paramecium
37	249.5	13.0	280	2	069808	069808 streptomyce
38	248	12.9	292	2	09ABU5	09ABU5 caulobacter
39	246.5	12.8	280	2	069939	069939 streptomyce
40	241	12.5	220	10	09ZQH4	09ZQH4 arabidopsis
41	240	12.5	292	2	025452	025452 helicobacte
42	239	12.4	292	2	09ZL86	09ZL86 helicobacte
43	238.5	12.4	262	2	09L104	09L104 streptomyce
44	233.5	12.2	271	2	09CB47	09CB47 mycobacteri
45	230.5	12.0	294	2	09ZMC7	09ZMC7 helicobacte

## ALIGNMENTS

RESULT 1  
ID 076091 PRELIMINARY; PRT; 327 AA.  
AC 076091;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE NITRILASE HOMOLOG 1.  
GN NIT1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98337986; Pubmed=9671749;  
RA Pekarsky Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,  
RA Tiliab S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,  
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;  
RT "Nitrihlase and Fhl1 homologs are encoded as fusion proteins in  
RT Drosophila melanogaster and Caenorhabditis elegans";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
DR EMBL: AF069984; AAC39901.1;  
DR EMBL: AF069987; AAC39907.1;  
DR InterPro: IPR001110; UPF0012.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS01227; UPF0012; 1.  
SQ SQUENCE 327 AA; 35896 MW; 90F7FB9D4BA627B1 CRC64;

Query Match 90.7%; Score 1743; DB 4; Length 327;  
Best Local Similarity 100.0%; Pred. No. 4.1e-147;  
Matches 326; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 37 LGFTTRPHRFLSLICGLRIPQSLVCAQPRPRAMAISSSSCLPLVAVQVSTPDQ 96  
DB 2 LGFTTRPHRFLSLICGLRIPQSLVCAQPRPRAMAISSSSCLPLVAVQVSTPDQ 61  
QY 97 QNFETCAELVREARLACLAFLPEAFDFIARDPAETLHLSPEPLGKLLSEYTLARECG 156  
DB 62 QNFETCAELVREARLACLAFLPEAFDFIARDPAETLHLSPEPLGKLLSEYTLARECG 121

```
OY 157 LMLSIGGFHERGQDWEQOTKIYNCHVLLNSKGAVVATYRKTHLCDVEIFGQGMCESNST 216
      |||||||
Db 122 LMLSIGGFHERGQDWEQOTKIYNCHVLLNSKGAVVATYRKTHLCDVEIFGQGMCESNST 181

OY 217 MGPSLESVSPACKIGTGLAVCYDMRFPPLSLALQAAGAEILTYPSAFSGTGPAPMEVL 276
      |||||||
Db 182 MGPSLESVSPACKIGTGLAVCYDMRFPPLSLALQAAGAEILTYPSAFSGTGPAPMEVL 241

OY 277 LRAALETQCYVVAACGCRHHEKRASYGSHMVDPMGTVVVARCSGPGCLARIDLNTL 336
      |||||||
Db 242 LRAALETQCYVVAACGCRHHEKRASYGSHMVDPMGTVVVARCSGPGCLARIDLNTL 301

OY 337 ROLRRHLPVFOHRRPDLVGNLGHPLS 362
      |||||||
Db 302 ROLRRHLPVFOHRRPDLVGNLGHPLS 327
```

```
RESULT 2
O88526 PRELIMINARY: PRT: 323 AA.
ID O88526;
AC O88526;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE NITRILASE HOMOLOG 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP MEDLINE=98337986; PubMed=9671749;
RX Pekarasy Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Fhl1 homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL: AF069985; AAC40184.1;
DR MGD: MGI:1350916; NITL.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA: 35717 MW: 10151CEB151DF2C7 CRC64;
```

```
Query Match 77.1%; Score 1482; DB 11; Length 323;
Best Local Similarity 84.4%; Pred. No. 6.8e-124;
Matches 275; Conservative 27; Mismatches 20; Indels 4; Gaps 2;
```

```
OY 37 LGFTRPPRHFLSLCPGLRIPQSLVLCQOPPRAMAISSSCCELPVAVCVOTSTPDKO 96
      |||||||
Db 2 LGFTRPPRHQ--LLCTGRLRLTLPVLCQOPPRRTMS-SSTSWELPLVAVCVOTSTPDKO 57

OY 97 QNFKTCALVREARLGAACLAFLPEAFDFIARDPAETLHLSPLGKLEETQLARECG 156
      |||||||
Db 58 ENFKTCALVQEARLGAACLAFLPEAFDFIARNPAETLLSEPLNDLLGQTSOLARECG 117

OY 157 LMLSIGGFHERGQDWEQOTKIYNCHVLLNSKGAVVATYRKTHLCDVEIFGQGMCESNST 216
      |||||||
Db 118 LMLSIGGFHERGQDWEQONKIYNCHVLLNSKGSVVASYRKTHLCDVEIFGQGMRESNYT 177

OY 217 MGPSLESVSPACKIGTGLAVCYDMRFPPLSLALQAAGAEILTYPSAFSGTGPAPMEVL 276
      |||||||
Db 178 KPGTLEPPVKTTPAGKVGALICYDMRFPPLSLKLAQGAELITYPASFGSVGPAPMEVL 237

OY 277 LRAALETQCYVVAACGCRHHEKRASYGSHMVDPMGTVVVARCSGPGCLARIDLNTL 336
      |||||||
Db 238 LRAALETQCYVVAACGCRHHEKRASYGSHMVDPMGTVVVARCSGPGCLARIDLNTL 297
```

```
OY 337 ROLRRHLPVFOHRRPDLVGNLGHPLS 362
      |||||||
Db 298 QOMRQHLPVFOHRRPDLVGSIGHPLS 323
```

```
RESULT 3
O9R1N4 PRELIMINARY: PRT: 323 AA.
ID O9R1N4;
AC O9R1N4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98337986; PubMed=9671749;
RA Pekarasy Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Fhl1 homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL: AF069988; AAC40185.1;
DR MGD: MGI:1350916; NITL.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA: 35705 MW: F8CD7730713665EF CRC64;
```

```
Query Match 77.0%; Score 1479; DB 11; Length 323;
Best Local Similarity 84.4%; Pred. No. 1.3e-123;
Matches 275; Conservative 26; Mismatches 21; Indels 4; Gaps 2;
```

```
OY 37 LGFTRPPRHFLSLCPGLRIPQSLVLCQOPPRAMAISSSCCELPVAVCVOTSTPDKO 96
      |||||||
Db 2 LGFTRPPRHQ--LLCTGRLRLTLPVLCQOPPRRTMS-SSTSWELPLVAVCVOTSTPDKO 57

OY 97 QNFKTCALVREARLGAACLAFLPEAFDFIARDPAETLHLSPLGKLEETQLARECG 156
      |||||||
Db 58 ENFKTCALVQEARLGAACLAFLPEAFDFIARNPAETLLSEPLNDLLGQTSOLARECG 117

OY 157 LMLSIGGFHERGQDWEQOTKIYNCHVLLNSKGAVVATYRKTHLCDVEIFGQGMCESNST 216
      |||||||
Db 118 LMLSIGGFHERGQDWEQONKIYNCHVLLNSKGSVVASYRKTHLCDVEIFGQGMRESNYT 177

OY 217 MGPSLESVSPACKIGTGLAVCYDMRFPPLSLALQAAGAEILTYPSAFSGTGPAPMEVL 276
      |||||||
Db 178 KPGTLEPPVKTTPAGKVGALICYDMRFPPLSLKLAQGAELITYPASFGSVGPAPMEVL 237

OY 277 LRAALETQCYVVAACGCRHHEKRASYGSHMVDPMGTVVVARCSGPGCLARIDLNTL 336
      |||||||
Db 238 LRAALETQCYVVAACGCRHHEKRASYGSHMVDPMGTVVVARCSGPGCLARIDLNTL 297

OY 337 ROLRRHLPVFOHRRPDLVGNLGHPLS 362
      |||||||
Db 298 QOMRQHLPVFOHRRPDLVGSIGHPLS 323
```

```
RESULT 4
O9DBF2 PRELIMINARY: PRT: 290 AA.
ID O9DBF2;
AC O9DBF2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
```



```

GN NIT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochila H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staahl F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Cusack-Stinch S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK004988; BAB23723.1;
DR MGD: MGI:1350916; Nf1l.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
SQ SEQUENCE 290 AA; 31886 MW; 34493DBFF2170C71 CRC64;

Query Match 71.48; Score 1372; DB 11; Length 290;
Best Local Similarity 87.5%; Pred. No. 3.6e-114;
Matches 252; Conservative 22; Mismatches 14; Indels 0; Gaps 0;

QY 75 SSSCEPLVAVCOVSTPPKQONFKTCALVREARLGAFLPAFPFIARDPPTL 134
DB 3 SSTSMELPLVAVCOVSTPPKQONFKTCALVREARLGAFLPAFPFIARDPPTL 62
QY 135 HLEPEPLGKLEETYLARCGMLSLGFGHERGQDWEOTOKIYNCHVLLNSKGA VVATY 194
DB 63 LLEPEPLGDLGGYSLARCGIMLSLGFHERGQDWEONQKTYNCHVLLNSKGS VVASY 122
QY 195 RKTHLCDVEIPGGCPMCESTYMPGSLSPVSTPACRTGLAVCYDMKFPFLSLALAOAG 254
DB 123 RKTHLCDVEIPGGCPMCESTYMPGSLSPVSTPACRTGLAVCYDMKFPFLSLALAOAG 182
QY 255 AEILTPSASGISTGPAMHVEVLLARAIETOCYVVAACGGRHHEKASVGHSHVVPWG 314
DB 183 AEILTPSASGISTGPAMHVEVLLARAIETOCYVVAACGGRHHEKASVGHSHVVPWG 242
QY 315 TVVARCEGPGCLARIDLNYLRQLRRHLPVQHRRPDLVGSGLHPLS 362
DB 243 TVVARCEGPGCLARIDLNYLRQLRRHLPVQHRRPDLVGSGLHPLS 290

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC PubMed=10959838;
RX Pace H.C., Hodawadekar S.C., Draganescu A., Huang J., Bieganski P.,
RA Pekarsky Y., Croce C.M., Brenner C.,
RT "Crystal structure of the worm NifH protein Rosetta Stone protein reveals a
RT Nif tetramer binding two Fe-S clusters."
RL Curr. Biol. 10:907-917(2000).
DR EMBL: AF284575; AAF87104.1;
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
SQ SEQUENCE 288 AA; 32061 MW; B3673A80609BF042 CRC64;

Query Match 49.88; Score 956; DB 13; Length 288;
Best Local Similarity 62.7%; Pred. No. 3.7e-77;
Matches 173; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 82 PLVAVCOVSTPPKQONFKTCALVREARLGAFLPAFPFIARDPPTLHLEPLIG 141
DB 7 PLVAVCOVSTPPKQONFKTCALVREARLGAFLPAFPFIARDPPTLHLEPLIG 66
QY 142 KLEPEYTLARECGMLSLGFGHERGQDWEOTOKIYNCHVLLNSKGA VVATYRTHLCD 201
DB 67 GDTIORTYTLARECGMLSLGFGHERGQDWEOTOKIYNCHVLLNSKGA VVATYRTHLCD 126
QY 202 VEIRGQCPMCESTYMPGSLSPVSTPACRTGLAVCYDMKFPFLSLALAOAGHLEPLYP 261
DB 127 VDLONGVSLRSESSSLPGLAELIRPITSPAGRTGLAVCYDMKFPFLSLALAOAGHLEPLYP 186
QY 262 SAFGSIPTPAHVEVLLARAIETOCYVVAACGGRHHEKASVGHSHVVPWGVARCS 321
DB 187 SAFGSIPTPAHVEVLLARAIETOCYVVAACGGRHHEKASVGHSHVVPWGVARCS 246
QY 322 EGPGLCLARIDLNYLRQLRRHLPVQHRRPDLVGSGLHPLS 357
DB 247 EGPGLCLARIDLNYLRQLRRHLPVQHRRPDLVGSGLHPLS 282

```

```

RESULT 6
ID 076464 PRELIMINARY; PRT; 460 AA.
AC 076464;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NITFAIT PROTEIN.
GN NITFAIT OR CG7067.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Adair J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,

```

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hoston D., Houston K.A., Howland T.J., Mei M.-H., Ibeigbam C.,  
RA Jastli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacible J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Putl V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Splier E., Spridling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-98337986; PubMed-9671749;  
RA Pekarsky Y., Campigillo M., Siprasvill Z., Druck T., Sedkov Y.,  
RA Tilib S., Draganesu A., Wermtuth P., Rothman J.H., Huebner K.,  
RT Buchberg A.M., Mazo A., Brenner C., Croce C.M.;  
RT "Nitrilase and Fhit homologs are encoded as fusion proteins in  
Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
DR EMBL: AE003467; AAF47347.1; -;  
DR EMBL: AF069989; AAC39137.1; -;  
DR HSSP: P49789; 2F1T.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR InterPro: IPR003010; HIT.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR Pfam: PF01230; HIT; 1.  
DR PROSITE: PS00892; HIT; 1.  
SQ SEQUENCE 460 AA; 52231 MW; 81121A00BC37706 CRC64;

Query Match 37.0%; Score 710.5; DB 5; Length 460;  
Best Local Similarity 48.6%; Pred. No. 4,5e-55;  
Matches 143; Conservative 47; Mismatches 97; Indels 7; Gaps 3;  
OY 62 VTCADPRPAMISSSSCLPLVAVCOVSTPDKOONFKTCALVDEARLGAFLPDE 121  
DB 13 VAIHQQLERMSVQKRKDSATLAVGOMRSTSDKANLSQVLELVDRASQNMCLFDE 72  
OY 122 ADFDIARDAETLHSEPLGKLLSEYTOARCEGLMSLGFHFGHGOQWEOQKLYNCH 181  
DB 73 CCDFVVESTQTILESEGLDGLMAOYRELAKCNKTIWISLGVHENRND-----OKFNAH 127  
OY 182 VLNSKGAVVATYRKTLDVETPGGPMCSNSTMPGSPLESVPSTPAKGTGLAVCYDM 241  
DB 128 VLINKEGELAAVYRKLMHEDVTK-EVRLRESDTVPKGCLERPVSTPGVIGLQICYDL 186  
OY 242 REPPELSLAQAGAEILTPSAFGSTTTPAHMEVLLRAAIEITQCYVVAACGGRHNEK 301  
DB 187 RAEPRAVLLRKLGANLLTPSAFTYATGKAHWEILLRAAIEITQCYVVAACGGRHNEK 246  
OY 302 ASYGSMMVDPGTVVAVRCSEGP-GICLARIDNLYRLRHLPVQHRPDLX 354  
DB 247 OSMGSHMIVSPGNNVLADCSBDELIGTAEVDLSVLQSLYQTPCFEHRNDIY 300

RESULT 7  
ID 076463 PRELIMINARY; PRT; 440 AA.  
AC 076463;  
DT 01-NOV-1998 (TREMblrel. 08, Created)  
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE NITRILASE AND FRAGILE HISTIDINE TRIAD FUSION PROTEIN NITFIHT.  
GN NITFIHT OR Y56A3A.13.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-98337986; PubMed-9671749;  
RA Pekarsky Y., Campigillo M., Siprasvill Z., Druck T., Sedkov Y.,  
RA Tilib S., Draganesu A., Wermtuth P., Rothman J.H., Huebner K.,  
RT Buchberg A.M., Mazo A., Brenner C., Croce C.M.;  
RT "Nitrilase and Fhit homologs are encoded as fusion proteins in  
Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99069613; PubMed-9851916;  
RA Mathews L.;  
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99069613;  
RA none;  
RT "Genome sequence of the nematode *C. elegans*: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL: AF069986; AAC39136.1; -;  
DR EMBL: AL132860; CAB60517.1; -;  
DR HSSP: P49789; 2F1T.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR InterPro: IPR001310; HIT.  
DR InterPro: IPR001110; UPF0012.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR Pfam: PF01230; HIT; 1.  
DR PROSITE: PS00892; HIT; 1.  
DR PROSITE: PS01227; UPF0012; 1.  
SQ SEQUENCE 440 AA; 49936 MW; 34EBD223CE92975 CRC64;

Query Match 36.5%; Score 702; DB 5; Length 440;  
Best Local Similarity 50.2%; Pred. No. 2,4e-54;  
Matches 136; Conservative 43; Mismatches 90; Indels 2; Gaps 1;  
OY 84 VAVCOVSTPDKOONFKTCALVDEARLGAFLPDEAFDTPARDAETLHSEPLGK 143  
DB 17 IAVCOVSTPDKOONFKTCALVDEARLGAFLPDEAFDTPARDAETLHSEPLGK 76  
OY 144 LEEYTOARCEGLMSLGFHFGHGOQWEOQKLYNCHVLNSKGAVVATYRKTLDV 203  
DB 77 YMEKYRELARKHNIWISLGLHNK--DPSDAHPWNTIILIDSDVTRAEVNKLHLEFDE 134  
OY 204 IPGQGMCSNSTMPGSPLESVPSTPAKGTGLAVCYDMREPPELSLAQAGAEILTPSA 263  
DB 135 IPGKVMLESSEFSKATEMI PVDPDTIGRIGLSTCYDVERPELSLNNRRGQALLSFP 194  
OY 264 FGSITPAHMEVLLRAAIEITQCYVVAACGGRHNEKRSYGSVVDPMGTVVAVRCSE 323  
DB 195 FTNLNGLAMETILRAAIEITQCYVVAACGGRHNEKRSYGSVVDPMGTVVAVRCSE 254  
OY 324 PGICLARIDNLYRLRHLPVQHRPDLX 354  
DB 255 VDMCFEALDLSYVDTLREMOVPFSHRSDLY 285  
RESULT 8

09LE50  
ID 09LE50 PRELIMINARY: PRT: 316 AA.  
AC 09LE50:  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE NITRILASE 1 LIKE PROTEIN.  
GN ATAG08790.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McCombie R.W., Spiegel L.A., Huang E.N., Nascimento L.U.,  
RA de la Bastide M., Vill D.M., Preston R.R., Matero A., Shah R.,  
RA O'Shaughnessy A., Rodriguez M., Shekher M., Schutz K., See L.H.,  
RA Swaby I., Habermann K., Dedhia N.N., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vill D.M.,  
RA Shekher M., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,  
RA Preston R.R., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL161813; CAB82115.1; -  
DR EMBL; AL161512; CAB78004.1; -  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
SQ SEQUENCE 316 AA: 34807 MW: 1D740F48DF9A202C CRC64;

Query Match 33.6%; Score 645.5; DB 10; Length 316;  
Best Local Similarity 44.2%; Pred. No. 1.7e-49;  
Matches 129; Conservative 53; Mismatches 91; Indels 19; Gaps 5;

QY 68 RRRAAIASSSCGLP-----VAVCOVSTPDKQONFKTCAELVREARIGACL 116  
DB 12 KSLFTTRITLSSQIPLTMAITVKNKTVRAAQMSTVNDMTNATCSRLVQEAALAGAKL 71  
QY 117 AFLPEAFDFIARDPAETHTLSEPIGLKLEEYTLARECGIMLGLGPFHRRGDMWQOTK 176  
DB 72 ICFPEHFVSFGKEGESVKAIEPLDGPWERICSLARDNSIMLSLGGFQDERFDD----TH 127  
QY 177 IYNCHVLNLSKAGVAVATYKTHLCDEIPIGOGPCESSNSTMPGPSLESPVSPAGRIGLA 236  
DB 128 LCNFTVHVIDDAMGIRDTYQKMLFDVDVPGSSSKESSTFVPGTKIVS--VDSVGRGLGL 186  
QY 237 VCYDMRFPPLSLALA-QAAGELITYSAGSTIGPAHMEVILRAKALETCQYVAAAQCG 295  
DB 187 VCYDLRFPKITYOOLRFEORAQVLLVPSAFTKYVGEAHMELLARALETCQYVAAAQAG 246  
QY 296 RHHRKRASGSHMVDVPGTVAARCE--GPGICLARIDILNTYRRLHLPV 345  
DB 247 KNEKRRESIGDTLLIDPMGTIVGRPLDRVSTGIIVADIDFSLIDSVRTKMPI 298

RESULT 9  
AC 094660: PRELIMINARY: PRT: 276 AA.  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE HYPOTHEICAL 30.4 KDA NITRILASE-LIKE PROTEIN C651.02 IN CHROMOSOME II.  
GN SPBC651.02.

OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-972;  
RC Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO THE N-TERMINAL OF NITRILASES.  
DR EMBL; AL035570; CAB37598.1; -  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
KW Hypothetical protein; Lyase.  
FT ACT\_SITE 156  
SQ SEQUENCE 276 AA: 30421 MW: 07AA741A54297E07 CRC64;

Query Match 30.0%; Score 575.5; DB 3; Length 276;  
Best Local Similarity 44.2%; Pred. No. 2.5e-43;  
Matches 126; Conservative 50; Mismatches 86; Indels 23; Gaps 8;

QY 81 LPLVAVCOVSTPDKQONFKTCAELVREARIGACLPEAFDFIARDPAETLHT-SEP 139  
DB 1 MTLAAVAQGLNSSGSLIKNLALICKELISQAARAKACIFFPEADDFIAHNDDEALIELTNHP 60  
QY 140 LGGKLEEYTLARECGIMLGLGPFHRRGDMWQOTKYNCHVLNLSK-----GAVVA 192  
DB 61 DCSKTRFDVRESATKSHISFVNI--CVHE-----PSKVN--KLNSLFIPELHGIIIS 110  
QY 193 TYKTHLCDEIPIGOGPCESSNSTMPGPSLESPVSPAGRIGLAICYDMRFPPLSLALAQ 252  
DB 111 RYSKAHLFDEIINGPTLKESNTLREGAILPCKKPLGKVGASLCPDIPPDQAIKRN 170  
QY 253 AGAEILTPYSPAFSGTIGPAHMEVILRAKALETCQYVAAAQCGRHHRKRASGSHMVDV 312  
DB 171 MGAHITTPSAFTKTKGAHMEVILRAKALDSQCYVAPAAQGGKHNKRASGSHMIVDP 230  
QY 313 MGTVAARCE--GP-GICLARIDILNTYRRLHLPVQHRRPDLX 354  
DB 231 MGVIVAIQVSDISSPNGLIFADLDLNLVDVHYRTYPLL--RRNDLY 273

RESULT 10  
ID 027839 PRELIMINARY: PRT: 272 AA.  
AC 027839:  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE N-CARBAOYL-D-AMINO ACID AMIDOHYDROLASE.  
GN MTH1811.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OC Methanobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-DELTA H;  
RC MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Polther B., Qiu D.,  
RA Spadafora R., Vicare A., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.T., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
RT Complete genome sequence of Methanobacterium thermoautotrophicum  
RT delta: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL; AE000934; AAB86277.1; -  
DR InterPro: IPR003010; CN\_hydrolase.



Db 224 VVMPWGEVLAKAGTEEAIVYSDIDLKLAETLRQIIPVFRQKRSPLY 269

RESULT 13  
ID 083040 PRELIMINARY; PRT; 270 AA.  
AC 083040;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE ORF270 PROTEIN.  
OS Plectonema boryanum.  
OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.  
OX NCBI\_Taxid=1184;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D65;  
RA Matsumura T., Fujita Y., Hase T.;  
RT "Cyanobacterial ferredoxin gene."  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB017194; BA32602.1;  
DR InterPro: IPR001110; UPF0012.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS01227; UPF0012; 1.  
SQ SEQUENCE 270 AA; 29639 MW; 131C447C00BF740 CRC64;

Query Match 24.9%; Score 479; DB 2; Length 270;  
Best Local Similarity 39.6%; Pred. No. 9.3e-35;  
Matches 112; Conservative 40; Mismatches 95; Indels 36; Gaps 7;

QY 83 LVAVQVTSPPKQNFKAELVREARUGACLAFLPEAFDFTARDPAETLHLSEPLG 142  
DB 5 LAAVQVNSPDLQKLNLAQAELEIIVRRCAELIGLPENFSFGDEE-----G 53  
QY 143 KLEETVQLREGSLMS-----LGGFH---ERQGDWQOTQKIVNCHVLLNSKG 188  
DB 54 KIAQA-DEIVNESKEFLKTAQAQVQVITLGGYVPEVPG-----KVNTALLVAPNG 105  
QY 189 AVVATYKTHLCEVIEIPGQPMCESNSTMPSPLESFV--STPAKIGLAVCYDMRPPEL 246  
DB 106 ELARVRYKHLFDVNLDPGNTYRESATVAVSTR-PRTHPKDGLNGLSICYVRPEL 164  
QY 247 SLAQAQAEILYPSAFSGSTGPAHWEVLARALETQCYVAAACGRHHEKRAYGH 306  
DB 165 YRQMTQGAELVFPAAFTAHGKDHQVLLQARAIENTCYVIAPAGTGRHNSLSRSHGH 224  
QY 307 SMVVDPMGTAVVARGSEGPLARIDNLRLRLHPVFOHR 349  
DB 225 AMIDPMGVILADAGEMTGAIAELAPGRLEQVRROMPSLOHR 267

RESULT 14  
ID 09JHM2 PRELIMINARY; PRT; 276 AA.  
AC 09JHM2;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE NIT PROTEIN 2 (1190017B19RIK PROTEIN).  
GN NIT2 OR 1190017B19RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Pubmed=10959838;  
RA Pace H.C., Hodawadekar S.C., Dragescu A., Huang J., Bleganowski P.,  
RA Pekarsky Y., Croce C.M., Brenner C.;  
RT "Crystal structure of the worm Nitfhl Rosetta Stone protein reveals a

RT Nit tetramer binding two Fhl dimers.";  
RL Curr. Biol. 10:907-917(2000).

RESULT 15  
ID 09KUD4 PRELIMINARY; PRT; 275 AA.  
AC 09KUD4;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE HYPOTHETICAL PROTEIN VC0421.  
GN VC0421.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_Taxid=666;  
RN [1]  
RP SEQUENCE FROM N.A.

Query Match 24.7%; Score 474.5; DB 11; Length 276;  
Best Local Similarity 37.2%; Pred. No. 2.4e-34;  
Matches 103; Conservative 52; Mismatches 103; Indels 19; Gaps 6;

QY 84 VAVQVTSPPKQNFKAELVREARUGACLAFLPEAFDFTARDPAETLHLSEPL 140  
DB 6 LALIQLVSSIKSDNLTFRACSLVREARUGANIVSLPCFN---SPGYTFYFPDYAKI 61  
QY 141 GKLLEETVQLREGSLMS-----LGGFH---ERQGDWQOTQKIVNCHVLLNSKGAVATYKTHLC 200  
DB 62 PESTQKLEEVAKESSTIYLIGSGIPE-----EDACKLYNTCSVGPDSGLYKRRKHLF 116  
QY 201 DVEIPGQPMCESNSTMPSPLESFVSTPAKIGLAVCYDMRPPELSTALAQAQAEILTY 260  
DB 117 DIDVPGKITFQESKTLSPGDSF-STFDTPYCKVIGICIDYDMFAELAQIYAQRGOLLVY 175  
QY 261 PSAFSGSTGPAHWEVLARALETQCYVAAACGRHHEKRAY--CHSNVVDPMGTAVV 317  
DB 176 PGAFNLTTGPAHWEVLQARAIVDQVYVATASPA---RDDRASYVAMGHSVVDPMGQVL 232  
QY 318 ARCSGPGCLARIDNLRLRLHPVFOHRPPLY 354  
DB 233 TRAGTEETILYSDIDLKLAETLRQIIPILKORRADLY 269



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 17:02:50 ; Search time 21.41 Seconds

(without alignments)  
619.928 Million cell updates/sec

Title: US-09-357-675C-25

Perfect score: 1 PLAAAXLAPDRPRDRLRMV.....LPVFQHRPRDLGNLGHPLS 362

Sequence:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	519	27.0	307	1 YJM6_YEAST	P47016 saccharomyc
2	490	25.5	372	1 P55175	saccharomyc
3	448.5	23.3	322	1 YAU8_SCHPO	Q10166 schizosacch
4	420	21.9	291	1 YL85_YEAST	P49954 saccharomyc
5	356	18.5	262	1 YBEM_ECOLI	P39874 escherichia
6	353	18.4	262	1 YBEM_ECO57	P58054 escherichia
7	262.5	13.7	340	1 Y480_MYCTU	Q11446 pseudomonas
8	254.5	13.2	285	1 YPOO_PSEFL	P55176 pseudomonas
9	231	12.0	234	1 YAG5_STALU	P55178 staphylococ
10	221.5	11.5	346	1 NRJ3_ARATH	P46010 arabidopsis
11	220.5	11.5	261	1 YAG5_STRAU	P55177 staphylococ
12	214.5	11.2	346	1 NRJ1_ARATH	P32961 arabidopsis
13	198.5	10.3	339	1 NRJ2_ARATH	P32962 arabidopsis
14	195	10.2	346	1 ALAM_PSEAE	P11436 pseudomonas
15	179.5	9.3	393	1 BUP_RAT	Q03248 rattus norv
16	178.5	9.3	355	1 NRJ4_ARATH	P46011 arabidopsis
17	159	8.3	349	1 NRJ4_TOBAC	Q42965 nicotiana t
18	135	8.1	513	1 YHCK_BACSU	P54608 bacillus su
19	142	7.4	256	1 YAFV_ECOLI	Q47679 escherichia
20	142	7.4	345	1 ALAM_RHOER	Q01360 rhodococcus
21	131	6.8	576	1 NAE2_THEMA	Q90Y00 thermotoga
22	115.5	6.0	199	1 YI04_YEAST	P40447 saccharomyc
23	114.5	6.0	679	1 MADE_MYCTU	P71911 mycobacteri
24	113.5	5.9	356	1 NRJ4_ALCEA	P20960 alcaligenes
25	112.5	5.9	368	1 CYRH_GLOSO	P32964 glaucocercos
26	108.5	5.6	552	1 MADE_RHOCA	Q03638 rhodobacter
27	102	5.3	383	1 NRJ1_RHOER	Q02068 rhodococcus
28	102	5.3	567	1 MADE_AOUAE	Q67091 aquifex aeo
29	100	5.2	365	1 NRJ2_RHOER	Q03217 rhodococcus
30	98.5	5.1	511	1 LNT_PSEAE	Q92186 pseudomonas
31	97.5	5.1	680	1 MADE_MYCLE	Q9cb66 mycobacteri
32	96	5.0	794	1 SEIL_HUMAN	Q9ubv2 homo sapien
33	94.5	4.9	433	1 FUT4_RAT	P62994 rattus norv

34	94	4.9	542	1 LNT_CHLTR	O84539 chlamydia t
35	93	4.8	349	1 NRJ6_KLEPO	P10045 klebsiella
36	93	4.8	519	1 LNT_SYNY3	P74055 synechocyst
37	92.5	4.8	520	1 VNN2_HUMAN	Q95498 homo sapien
38	91.5	4.7	541	1 LNT_RHIME	Q52910 rhizobium m
39	91	4.7	512	1 VNN1_MOUSE	Q92018 mus musculu
40	90	4.7	439	1 LNT_AOUAE	Q67000 aquifex aeo
41	90	4.7	2209	1 P0IG_POLIS	P03301 poliovirus
42	89.5	4.7	642	1 HEM1_RAT	P13195 rattus norv
43	89.5	4.7	1257	1 ERB2_RAT	P06454 rattus norv
44	89	4.6	1255	1 ERB2_HUMAN	P04626 homo sapien
45	88	4.6	790	1 SEIL_MOUSE	Q92296 mus musculu

## ALIGNMENTS

RESULT 1	ID	YJM6_YEAST	STANDARD	PRT	307 AA.
AC	P47016				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DE	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	HYPOTHETICAL 34.7 KDA PROTEIN IN SPT10-GCD14 INTERGENIC REGION.				
CN	YJL126W OR J0706.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_Taxid=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRATIN-S288C / FY1679.				
RX	MEDLINE=97103775; PubMed=8948101;				
RA	Czepluch C., Kordes E., Pujol A., Jauniaux J.-C.;				
RT	"Sequencing analysis of a 40.2 kb fragment of yeast chromosome X				
RT	reveals 19 open reading frames including URA2 (5' end), TRJ1, PBS2,				
RT	SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCR7, GZF3, two tRNA genes,				
RT	three remnant delta elements and a Ty4 transposon."				
RL	Yeast 12:1471-1474(1996).				
CC	- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	EMBL: 249401; CA89421.1; -.				
DR	SGD: S0003662; YJL126W.				
DR	InterPro: IPR003010; CN_hydrolase.				
DR	InterPro: IPR001110; UPF0012.				
DR	Pfam: PF00795; CN_hydrolase; 1.				
DR	PROSITE: PS01227; UPF0012.1.				
KW	Hypothetical protein.				
SQ	SEQUENCE 307 AA; 34693 MW; 48787C43B10A828E CRC64;				
Query Match	27.0%	Score 519;	DB 1;	Length 307;	
Best Local Similarity	37.6%	Pred. No. 6.9e-38;			
Matches 114;	Conservative 56;	Mismatches 105;	Indels 28;	Gaps 6;	
QY	80	ELPLVAVQVYTPDPKQNFCTCAELVREARLGCACLAFLPEARDFTARPATLHL---	136		
DB	4	KLRKVAQAQCSSADLNKLVKELISEAQKADAVFLPEASDYLSQNPILHSRYLAOK	63		
QY	137	SEPLGKLEBYTOLARECGILSLG-GFH--ERGGD-WEGTQKIYMCIVLNSKAVVA	192		
DB	64	SPKFRLOLSSITDLVBDNSRNDVSGIVHLPPSEODLLEQNDVRVNVLLYIDHEGKITQ	123		
QY	193	TYRTHLADVEIIPGGWMCESNTMPSPLESPIVSTAGKIGLAVCVDMRPEPLSLAQAQ	252		

```
DB 124 EKQKHLFDVDPVNGPILKESKSVOPGKAIPIDIESPLKLSAICQDIRPFESLTKRS 183
OY 253 AGAEILTPSAGSITGPALHWEILRLARAIETQCYYVAAAACGRH----- 297
DB 184 MGAELICFSAITKTGEGHWEILGRARAVDTQCYVLMGQGMHDLSPENWEKOSHMSA 243
OY 298 ---HEKRASYSGHSMVDPMGTVVARC---SEGPGLCLARIDINYLRLRRLHVPFQHRP 351
DB 244 LEKSSRESGSHMSVIDPMGKIITAHADPSTVGQILADLDRELLOEIRNKPMIMNQORD 303
OY 352 DLX 354
DB 304 DLF 306

RESULT 2
Y601_SYNY3 STANDARD; PRT; 272 AA.
ID Y601_SYNY3
AC P55175;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 30.2 KDA PROTEIN SLL0601.
GN SLL0601.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-96127529; PubMed-8590279;
RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1 SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D64002; BAA10370.1;
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 272 AA; 30191 MW; 7787BA9B2BE003A1 CRC64;
```

```
Query Match 25.5%; Score 490; DB 1; Length 272;
Best Local Similarity 40.2%; Pred. No. 2e-35;
Matches 111; Conservative 46; Mismatches 99; Indels 20; Gaps 7;
```

```
OY 33 LVAVQVSTPRKQNEFKTCAELVREARLGLACLAFLPAEDFIARDPAETLHLEPGL- 141
DB 5 LAAALQMTSPRLNTELOCAEBELIDLAIVROGAELVGLPENFAFLG---ETEKLEQATAI 61
OY 142 GKLEEVPTQ-LAREGGLMISLGGF-----HERGQDEQYQKTYNCHVILLNSKGAIVATVR 195
DB 52 ARAATEKFLDTMAQRQVYTLLAGFPFPPVAGEAG-----KAYNTATYLLAPNGQELARKH 114
OY 136 KTHLCLDVEILPGGPMCESNSTMPGSLSPV--STPAKIGLAVCYDMRFPETSLALAOA 253
DB 115 KYHLFDVNPVNDQNTYWESTVMAQGY-PPVYHSDSFGMLGLSICDVAFPELYRLSLNQ 173
OY 254 GAEILTPSAGSITGPALHWEILRLARAIETQCYYVAAAACGRHHEKRASYSGHSMVDPW 313
```

```
DB 174 GADVLEVPAAFTAYCKDKHMOVLLQARAIENTCYVLAPOQTSCHYERRHTGHAMIDPW 233
OY 314 GTVVARCSEGPGLCLARIDINYLRLRRLHVPFQHR 349
DB 234 GYILADAGEKPGLAIAETINPDRLKQYRQOMPLOHR 269

RESULT 3
YAU6_SCHPO STANDARD; PRT; 322 AA.
ID YAU6_SCHPO
AC Q10166;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 35.7 KDA PROTEIN C26A3.11 IN CHROMOSOME I.
GN SPAC26A3.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RA Mclean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z69240; CAA93234.1;
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein.
SQ SEQUENCE 322 AA; 35679 MW; EA6F39B160C7F49F CRC64;
```

```
Query Match 23.3%; Score 448.5; DB 1; Length 322;
Best Local Similarity 32.8%; Pred. No. 9.8e-32;
Matches 108; Conservative 63; Mismatches 123; Indels 35; Gaps 9;
```

```
OY 46 RFLSLCLPGLR--IFQLSVYCAQPPRRAMAISSSCCLP-----LVAVQVSTPRDKQ 97
DB 4 KRFGLVQKTRSFPSLNCYTR---NIMSVASSL-VKDRPARRIGLVQANKDKSE 59
OY 98 NFKTCAELVREARLGLACLAFLPAEDFIARDP-----AETLIHSEPLGKLEET 148
DB 60 NQQLRLKLVLEAKKGSNVIPLPIFN---SPYGTGFNOYAEPLEESSP---SYQAL 111
OY 149 TOLAREGGLMISLGGHGHGQDEQYQKTYNCHVILLNSKGAIVATVR 208
DB 112 SSMARDQTKYTLTGGSIPER-----KDGKLYNTAMFDPGSKLIANRKHLDLDIDIPGV 166
OY 209 PMCESNSTMPGSLSPVSTPAKIGLAVCYDMRFPETSLALQAGAEILTPSAGSIT 268
DB 167 SFRSDSLSPGAM-TWDTFEGKFGGLGICVDIRPELAMIAARNGCSVMYIPGAFNLST 225
OY 269 GRAHWEILRLARAIETQCYYVAAAACGRHHEKRASYSGHSMVDPMGTVVARCSEGPGLCL 328
DB 226 GLHWEILRLARAVNENMVACCAPARDMADYHSGHSTVVDPEGKVIATTDKPSIV 285
OY 329 ARIDINYLRLRRLHVPFQHRPDLXGNL 357
DB 286 ADIDPSVMSTARNSVPIYQRRFDVYSEV 314
```



```

RESULT 4
ID YBEM_ECOLI STANDARD; PRT; 262 AA.
AC P39874;
DT 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHEMETICAL PROTEIN YBEM.
GN YBEM OR B0625/B0626.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aliba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajinara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horituchi T.;
RT "A 7.18-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurid O., Lew H., Lin D.,
RA Nemat A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-51 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95075315; PubMed=7984109;
RA Yamanaoka K., Mitani T., Ogura T., Niki H., Hiraga S.;
RT "Cloning, sequencing, and characterization of multicopy suppressors
RT of a mub mutation in Escherichia coli."
RL Mol. Microbiol. 13:301-312(1994).
RN [5]
RP SEQUENCE OF 31-78 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Robinson K., O'Keefe T., Church G.M.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [6]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JAN-2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE K12 SEQUENCE
CC DIFFERS FROM THE WILD-TYPE SEQUENCE DUE TO A FRAMESHIFT IN
CC POSITION 67.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U01902; AAB67751.1; -.
CC SGD: S0004343; YLR351C.
CC InterPro: IPR003010; CN_hydrolase.
CC InterPro: IPR001110; UPF0012.
CC Pfam: PF00795; CN_hydrolase; 1.
CC PROSITE: PS01227; UPF0012; 1.
CC Hypothetical protein.
KM SEQUENCE 291 AA; 32549 MW; A81374412008827 CRC64;
SQ

```

```

ID YBEM_ECOLI STANDARD; PRT; 262 AA.
AC P39874;
DT 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHEMETICAL PROTEIN YBEM.
GN YBEM OR B0625/B0626.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aliba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajinara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horituchi T.;
RT "A 7.18-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurid O., Lew H., Lin D.,
RA Nemat A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-51 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95075315; PubMed=7984109;
RA Yamanaoka K., Mitani T., Ogura T., Niki H., Hiraga S.;
RT "Cloning, sequencing, and characterization of multicopy suppressors
RT of a mub mutation in Escherichia coli."
RL Mol. Microbiol. 13:301-312(1994).
RN [5]
RP SEQUENCE OF 31-78 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Robinson K., O'Keefe T., Church G.M.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [6]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JAN-2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE K12 SEQUENCE
CC DIFFERS FROM THE WILD-TYPE SEQUENCE DUE TO A FRAMESHIFT IN
CC POSITION 67.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AE000167; AAC73726.1; ALT_FRAME.
CC DR EMBL: AE000167; AAC73727.1; ALT_FRAME.
CC EMBL: D90703; BAA35268.1; ALT_FRAME.

```



RC STRAIN-CDC 1551 / Oshkosh:  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bisai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: 277162; CAB00941.1; -  
 DR EMBL: AE006951; AAK44721.1; -  
 DR TIGR: MT0498; -  
 DR Tuberculist; RV0480c; -  
 DR Interpro: IPR003010; CN\_hydrolase.  
 DR Interpro: IPR001110; UPF0012.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS01227; UPF0012; 1.  
 KW Hypothetical protein; Complete proteome.  
 KW DOMAIN 192 201 VAL-RICH.  
 SQ SEQUENCE 340 AA; 35195 MW; 541EAC9CD1D439B9 CRC64;

Query Match 13.7%; Score 262.5; DB 1; Length 340;  
 Best Local Similarity 28.4%; Pred. No. 1.5e-15;  
 Matches 89; Conservative 41; Mismatches 140; Indels 43; Gaps 10;

QY 53 PGRLIQLSLVCAOPRPRAAIISSSCSELPVAVCOVTPRDKQONFKTCAELVEARL 112  
 DB 42 PGSCARCCALVAGPRILARRI-----ALAIQIRSTDPANILQVLKYAGENATA 91  
 QY 113 GAGIAPLPEAFDFIARDPAETLHLEPLGKLEETQLARECGMLSLGCFHERGDWE 172  
 DB 92 GAGLVEPEA--TMCRLGVPLRQVAEPDGMANGVARIATEGIVIAEMFTTGGG-- 147  
 QY 173 QIQKINCHVLL--NSKGAIVATYRKTHLQDVLPQGGPMCESNSTMPGSLSPVSTPA 230  
 DB 148 ---RVNTNLTIAAGPTPNOPDAHYHKIHLHYDA----FGFTESRTVAPG--REPVVVV 196  
 QY 231 G--KIGLAVCYDMRPELSALAOAGELLTYSARGSTIGPA--HWEVLRLARALETQ 285  
 DB 197 DGVRVGLTVCYDIRFPALYTELARRGAQLTAVCASWS--GPGKLEQWTLIARRALDSM 254  
 QY 286 CYVVAAGC-----GRHHEKRASYGSHMVVDPMGTVVAVRCSEPGICLARIIDLNTL 336  
 DB 255 SYVAAAGADPDGARTGVGASSAAPTGVGSLVAPLGEVYVVASGTQPOLLDIVDNY 314  
 QY 337 RQLRHLPVFQHR 349  
 DB 315 AAARDRIAVLRNQ 327

RESULT 8  
 YPOO\_PSEFL STANDARD; PRT; 285 AA.  
 ID YPOO\_PSEFL  
 AC P55176;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL 31.2 KDA PROTEIN IN POOF 5' REGION (ORF2).  
 OS Pseudomonas fluorescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OC NCBI\_TaxID=294;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CHA0.  
 RX MEDLINE=96064397; PubMed=8526497;  
 RA Schneider U., Keel C., Defago G., Haas D.;  
 RT "Tns-directed cloning of ppg genes from Pseudomonas fluorescens CHA0;  
 RT mutational inactivation of the genes results in overproduction of the  
 RT antibiotic pyoluteorin.";  
 RL Appl. Environ. Microbiol. 61:3856-3864(1995).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X87299; CA60729.1; -  
 DR Interpro: IPR003010; CN\_hydrolase.  
 DR Interpro: IPR001110; UPF0012.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS01227; UPF0012; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 285 AA; 31163 MW; 68B7C64F38CBDECB CRC64;

Query Match 13.2%; Score 254.5; DB 1; Length 285;  
 Best Local Similarity 31.5%; Pred. No. 6.1e-15;  
 Matches 92; Conservative 43; Mismatches 126; Indels 31; Gaps 13;

QY 74 ISSSCSELPVAVCOVSTP-DKQONFKTCAELVEARLGAAGIAPLPEAFDIARDPAE 132  
 DB 14 LSVSGVTMRVALYQCPRLPDVAGNLRHQHMEAT--DADLVLPPEMLSGSYNIGLE 71  
 QY 133 TL-HLSEPLGKLEETQLARECGMLSLGGEHERGDWEQTKYNCHVLSKGANV 191  
 DB 72 AVGALAEADGPPSAQRITAAIAQAGTAT-LYGYERSYDGO---IYNVOLIDAQGR 126  
 QY 192 ATYKRTHLIC-DVEITPGGPMCESNSTMPGSLSP-VSTPAKIGLAVCYDMRPELSLA 249  
 DB 127 CNVKTHTLFGDLD-----HSMFSAGEDDPVLVLDQWKIGPLCYDIEPENARR 176  
 QY 250 LAQGAELITVPSAFSGITGPAHWEVLRLARAITQCYVVAAGCGRHHEKRASY-GHSM 308  
 DB 177 LALAGAEILIVPTA-NMIPYDFVADVTIRAFAPENOCYAVAYVCG--HEEOIRYCGOSS 233  
 QY 309 VDPWGTIVVAVRCSEPGICLARIIDLNTL--RQLRHLPVFQHRPDLXGL 357  
 DB 234 TAAPDGRIALAGIDELIITGIDRLQMGESRLNRYL---SDRRPELYDL 282

RESULT 9  
 YAG5\_STALU STANDARD; PRT; 234 AA.  
 ID YAG5\_STALU  
 AC P55178;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL PROTEIN IN AGR OPERON (ORF 5) (FRAGMENT).  
 OS Staphylococcus lugdunensis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Staphylococcus.  
 OC NCBI\_TaxID=28035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93366158; PubMed=8359673;  
 RA Vandenesch F., Projan S.J., Kreiswirth B., Etienne J., Novick R.P.;  
 RT "Agt-related sequences in Staphylococcus lugdunensis.";  
 RL FEMS Microbiol. Lett. 111:115-122(1993).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L13334; AAA17975.1; -  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR001110; UPF0012.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS01227; UPF0012; 1.  
 DR Hypothetical protein.  
 FT NON\_TER  
 KW  
 SQ SEQUENCE 234 AA: 26495 MW: C1CBADCA5E1389A7 CRC64;

Query Match 12.0%; Score 231; DB 1; Length 234;  
 Best Local Similarity 31.1%; Pred. No. 5,3e-13;  
 Matches 61; Conservative 27; Mismatches 64; Indels 44; Gaps 6;

QY 177 IYNCHVLSKGAIVVATYRTHLCDEIPGQPMCESNTPMPSLESPTSPAG----- 231  
 DB 64 IFNTAFADTKGTQVINOYDKMLV-----PMUDEPAFLTAGKNVPE 104  
 QY 232 -----RIGLAVCYDMRPEPSELALAOAGAEILTY-----PSAFSITGPAHMEVLT 279  
 DB 105 TFKLSNGVAVTQMICVDLRFPELRLPARSGATIAFYVQWMPARLN-----HWQVLKRA 159  
 QY 280 RALETQCYVVAACGCRHHEKRASY-GHSNVDPMCTVVARCSEGFGLCIARDLVYLRQ 338  
 DB 160 RALENMNVYVIGNGCG--YDGTKQVAGHSVAINPNCETIOLSTKELTFTVIDDAVEQ 217  
 QY 339 LRHRLPVPQHRREDLY 354  
 DB 218 ORKALPVPDSLPVPHLY 233

RESULT 10  
 NR13\_ARATH STANDARD; PRT; 346 AA.  
 ID NR13\_ARATH  
 AC P46010; 004909;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NITRILASE 3 (EC 3.5.5.1).  
 OS NIT3 OR AT3G44320 OR T10D17.110.  
 GN Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosidia II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=94294436; PubMed=8022831;  
 RA Bartel B., Fink G.R.;  
 RT "Differential regulation of an auxin-producing nitrilase gene family  
 in Arabidopsis thaliana.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=98145459; PubMed=9484465;  
 RA Hillebrand H., Bartling D., Weller E.W.;  
 RT "Structural analysis of the nlt2/nlt1/nlt3 gene cluster encoding  
 nitrilases, enzymes catalyzing the terminal activation step in indole-  
 acetic acid biosynthesis in Arabidopsis thaliana.";  
 RL Plant Mol. Biol. 36:89-99(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;

MEDLINE=21016720; PubMed=11130713;  
 RA Salimounat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,  
 RA Farnham B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Delany M., Boutry M., Grivell L.A., Maché R., Puidomenech P.,  
 RA De Simone V., Choisy N., Artiguenave F., Robert C., Brotier P.,  
 RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,  
 RA Schaefer M., Weiller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Wurmbach M., Drzonek H., Erle H., Jordan N., Bangert S.,  
 RA Vezeli A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,  
 RA Reichelt J., Scharfe M., Schoen O., Barques M., Terol J., Climent J.,  
 RA Navarro P., Collado C., Perez-Perez A., Ottemwalder B., Duchemin D.,  
 RA Cooke R., Laude M., Berger-Llauró C., Purnelle B., Masny D.,  
 RA de Haan M., Maarse A.C., Alcazar J.-P., Cottet A., Casacuberta E.,  
 RA Monfort A., Argilou A., Flores M., Liguori R., Vitale D.,  
 RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Wals A., Utterback T., Fujii C.T., Shea T.P.,  
 RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Millscher J., Sellers P., Gill J.E., Feldblum T.V.,  
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Frazer D., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 thaliana".  
 RL Nature 408:820-822(2000).  
 CC -I- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE  
 INDOLE-3-ACETIC ACID.  
 CC -I- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).  
 CC -I- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.  
 CC -I- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U09959; AAA19627.1; -  
 DR EMBL; Y07648; CAA68936.2; -  
 DR EMBL; AL353865; CAB89000.1; -  
 DR InterPro: IPR003010; CN\_hydrolase.  
 DR InterPro: IPR000132; Nitril\_cyn\_hydrolase.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
 KW Hydrolase; Multigene family; BY SIMILARITY.  
 FT ACT\_SITE 186  
 SQ SEQUENCE 346 AA: 38022 MW: 70CF421547F2B5E CRC64;

Query Match 11.5%; Score 221.5; DB 1; Length 346;  
 Best Local Similarity 27.4%; Pred. No. 5.6e-12;  
 Matches 86; Conservative 44; Mismatches 113; Indels 71; Gaps 14;

QY 91 STPDKOONFKTCAELVREARALGACLAFLPEAF-----DFTARDPAET 133  
 DB 41 ATLDKAEKF-----IVEAASKGAKLVLPPEAFIGYPRGFRGLAVGVHNEGRDEFNRN 94  
 QY 134 LHLSF-PLOGKLEETQTOLAREGGLSLGFGFERQDMEDQTKYNCHVLSKGAIVVA 192  
 DB 95 YHSAKIVPGPEVERLAELAGKNNVHLVNGALEKDG-----YTLCTALFFSPQGF 148  
 QY 193 TYRTHLCDE--IPGQPMCESNTPMPSLESPTSPAGKIGLAVCDMRPELSLAL 250  
 DB 149 KHKVMPSTSLERCITWGOG-----DGSTIP-----YDTPIGKIGALICENRNPPLYRTAL 198



```

RL Nature 408:820-822(2000).
RN [5]
RP CHARACTERIZATION.
RC STRAIN=CV, LANDSBERG, ERRECTA.
RA MEDLINE=94286570; PubMed=8016109;
RT Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;
RT "Molecular characterization of two cloned nitrilases from Arabidopsis
RT thaliana: key enzymes in biosynthesis of the plant hormone indole-3-
RT acetic acid."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
CC -!- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE
CC INDOLE-3-ACETIC ACID.
CC -!- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, BUT AT A
CC VERY LOW LEVEL DURING THE FRUITING STAGE.
CC -!- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X63445; CAA45041.1; -
DR EMBL: U38845; AAB05221.1; -
DR EMBL: Y07648; CAA68935.2; -
DR EMBL: AL353865; CAB86999.1; -
DR PIR: S22398; S22398.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR000132; Nitril_cyn_hydrolase.
DR Pfam: PF00795; CN_hydrolase_1.
DR PROSITE: PS00920; NITRIL_CHT_1; 1.
DR PROSITE: PS00921; NITRIL_CHT_2; 1.
DR KW Hydrolase; Multigene family.
FT ACT_SITE 186 BY SIMILARITY.
FT CONFLICT 312 Y -> H (IN REF. 2).
FT FT 312
SQ SEQUENCE 346 AA; 38178 MW; 8D4F8B7CAD1E3C1F CRC64;

```

```

AC P32962; Q96505;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-ANG-2001 (Rel. 40, Last annotation update)
DE NITRILASE 2 (EC 3.5.5.1).
GN NIT2 OR AT3G44300 OR T1D017-90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, LANDSBERG, ERRECTA; TISSUE=Leaf;
RX MEDLINE=94286570; PubMed=8016109;
RA Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;
RT "Molecular characterization of two cloned nitrilases from Arabidopsis
RT thaliana: key enzymes in biosynthesis of the plant hormone indole-3-
RT acetic acid."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=94294436; PubMed=8022831;
RA Bartel B., Fink G.R.;
RT "Differential regulation of an auxin-producing nitrilase gene family
RT in Arabidopsis thaliana."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Zhou L., Bartel B., Thornburg R.W.;
RT "Nucleotide sequence of a pathogen induced nitrilase gene from
RT Arabidopsis thaliana. Nit2."
RL (In) Plant Gene Register PKR96-006.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unselid M.,
RA Fartman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Maché R., Puigdomenech P.,
RA De Simone V., Cholsine N., Artiguenave F., Robert C., Brotier P.,
RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurzbach R., Drzonek H., Erle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Clement J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,
RA Cooke R., Landie M., Berger-Llauró C., Purnelle B., Masny D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argitrou A., Flores M., Liguori R., Vitale D.,
RA Manhaupt G., Haase D., Schoof H., Rüd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.K., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utecherback T., Fujii C.Y., Shea T.P.,
RA Cressy T.H., Haas B., Meiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltscher J., Sellers P., Gill J.E., Feldolym T.V.,
RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sakamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."
RL Nature 408:820-822(2000).
CC -!- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE
CC INDOLE-3-ACETIC ACID.
CC -!- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
CC -!- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.
CC -!- DEVELOPMENTAL STAGE: BARELY DETECTABLE IN YOUNG ROSETTES, BUT IS

```

CC STRONGLY EXPRESSED DURING BOLTING, FLOWERING, AND ESPECIALLY FRUIT  
CC DEVELOPMENT..

CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X68305; CAA48377.1; -  
DR EMBL: U09958; AAB60275.1; -  
DR EMBL: U38845; AAB05220.1; -  
DR EMBL: AL333865; CAB86986.1; -  
DR PIR: S31969; S31969.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR InterPro: IPR00132; Nitril\_cyn\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase.1.  
DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase; Multigene family.  
FT ACT\_SITE 179 179 BY SIMILARITY.  
FT CONFLICT 37 37 E -> G (IN REF. 3).  
FT CONFLICT 48 48 S -> T (IN REF. 3).  
SQ SEQUENCE 339 AA: 37153 MW: 06CDE76D2FDC24A7 CRC64:

Query Match 10.3%; Score 198.5; DB 1; Length 339;  
Best Local Similarity 25.9%; Pred. No. 5.5e-10;  
Matches 78; Conservative 47; Mismatches 115; Indels 61; Gaps 12;

QY 94 DKQNFKTCALVREARLACIALFLPEAF-----DPIARDPAETLHL 136  
DB 31 DHPATLEKANKRTIVEMASGSELVPEPAFICGYPFRGLGVGHNEGRDEFKRYHA 90  
QY 137 SE-PLGKLELEYTOLARCGMLSLGCFHERGQDWEQOKIYNCHVNLNSKAAVATYR 195  
DB 91 SAIKVPGEPEVLELAKGNVYLVVGALEKDG-----YTYLCTALFFSPQCGFLGKR 144  
QY 196 KTHLQDVE--IPQCGPMCSNSTMPSPLESPPSTPAGKIGLAVCYDMRPPELSLAQA 253  
DB 145 KLMPTSLERCINGCG-----DGSITP-----YDPTIGKGAICMENRPLRYTALYAK 194  
QY 254 GAEIITYPSEGSTGPAHWEVLLARALETQCYVVAQA-CGR-----HH 298  
DB 195 GIELCAPPADGS---KEMQSMLEHIAEGCGFVLACQFCLEKDFPDHPLYLFTDWD 250  
QY 299 EKR---ASYGHSWVDPMPTGVVARCS-EGPGICLARIDLNYLRQLRHL-PVFGHRRPD 352  
DB 251 DKPEPSIYSGGSVLIISPLGQVLAPNFESEGLITADLDGDVAAKLYFDVSGHYSRPD 310  
QY 353 L 353  
DB 311 V 311

RESULT 14  
ALAM\_PSEAE STANDARD: PRT: 346 AA.  
AC P11436:  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ALIPHATIC AMIDASE (EC 3.5.1.4).  
GN AME OR PA3366.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE.

RC STRAIN-PAC142;  
RX MEDLINE=87219101; PubMed=3108029;  
RA Ambler R.P., Anfret A.D., Clarke P.H.;  
RT "The amino acid sequence of the aliphatic amidase from Pseudomonas  
RT aeruginosa.";  
RL FEBS Lett. 215:285-290(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87219102; PubMed=3108030;  
RA Brammar W.J., Charles I.G., Matfield M., Liu C.-P., Drew R.E.,  
RA Clarke P.H.;  
RT "The nucleotide sequence of the amE gene of Pseudomonas aeruginosa.";  
RL FEBS Lett. 215:291-294(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AFCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Iori S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -1- FUNCTION: ALLOWS THE ORGANISM TO GROW ON SHORT CHAIN-LENGTH  
CC ALIPHATIC AMIDES.  
CC -1- FUNCTION: ENABLES THE ORGANISM TO USE ACETAMIDE AS BOTH CARBON  
CC AND A NITROGEN SOURCE.  
CC -1- CATALYTIC ACTIVITY: A MONOCARBOXYLIC ACID AMIDE + H(2)O = A  
CC MONOCARBOXYLATE + NH(3).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M27612; AAA25697.1; -  
DR EMBL: AE004759; AAG06754.1; -  
DR PIR: A26741; A26741.  
DR InterPro: IPR003010; CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase.1.  
KW Hydrolase; Complete proteome.  
FT CONFLICT 40 40 E -> D (IN REF. 1 AND 2).  
FT CONFLICT 169 169 G -> P (IN REF. 2).  
FT CONFLICT 317 317 E -> D (IN REF. 1 AND 2).  
SQ SEQUENCE 346 AA: 36495 MW: E19CEB474EB92B93 CRC64;

Query Match 10.2%; Score 195; DB 1; Length 346;  
Best Local Similarity 25.2%; Pred. No. 1.1e-09;  
Matches 76; Conservative 43; Mismatches 129; Indels 54; Gaps 11;

QY 76 SSSCEPLVAVCVSTPRKQNFKTCALVREARL-----GACIALFPE-A 122  
DB 7 SSSNDYGVAVVNV-----KMPRLHTAAEVLNARKIAEMIVGKQGLPGMDLVVPEYS 61  
QY 123 FDIARDPAETLHLSEPLGKLELEYTOLARCGMLSLGCFHERGQDWE--TQKLYNC 180  
DB 62 LOGIMDPAEMMETAAVAIPBEETEIFSRAOKRAVW--GVFSLTGRHHEHPKRAYNT 118  
QY 181 HVLNSGAVVATYRKTHLQDVEIPQCGPMCSNSTMPSPLESPPSTPAGKIGLAVCYD 240  
DB 119 LVLDNNGEIVQKTKIT-----PWCPIBGWIPGQOTYVSEPKMKKISLITICDD 168  
QY 241 MRPELSLAAQAGAEILT-----YPSAFGSTGPAHWEVLLARALETQCYVVAQAAC 294  
DB 169 GNPPEIWRDCAMKGAELIVCGGYMPAKDQV-----MMKAMAMANNVCVAVANAA 221

Qy 295 GRNHEKRYASGHWVDDMGTVVARCSGR - GICLARIIDLNYR ----- QLRNHLRVFQ 347  
 Db 222 G-FDDVGYGFHSHALIGDGTLTGECGEENGIGYADLSLQIFDAANQSQNHLEKIL 280

RESULT	15
BU_P_RAT	
ID	BU_P_RAT
	STANDARD;
	PRT;
	393 AA.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

Search completed: April 26, 2002, 17:02:51  
Job time: 454 sec

Query Match	9.3%	Score 179.5	DB 1	Length 393
Best Local Similarity	26.0%	Pred. No. 2.9e-08		
Matches 77: Conservative	41	Mismatches 139	Indels 39	Gaps 11





**THIS PAGE BLANK (USPTO)**



A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.

A:Reference number: A85001; MUID:20083488

A:Accession: D85088

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <STO>

A:Cross-references: GB:NC\_001268; NID:g7267521; PIDN:CAB78004.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g08790

A:Map position: 4

#### Query Match

33.6%; Score 645.5; DB 2; Length 316;

Best Local Similarity 44.2%; Pred. No. 1e-48;

Matches 129; Conservative 53; Mismatches 91; Indels 19; Gaps 5;

68 RPRMAISSSCEPL-----VAVQVSTPDKOONFKCALVREAAAGACL 116

12 KPSLETRITSSQIPLMATVNTKVRVAAAMQSVNDLMTFATCSRLVQFAALAGAKL 71

117 AFLPEAFDPIARDPAETLHLEPLGKLEETOLARECGMLSLGFERGODWEQOTOK 176

72 ICFPENEFSVDKGESEKIAEPIDGPMERYCSLARSNITWLSLGGQERDD-----TH 127

177 IYNCVILLNSKGAVALYATRYKTHLCDVEIPGQPMCESNSTMPGSPLESVPSTPAKIGLA 236

128 LCNTHVVIDDAGMIRDTYQKMHLEFDVDPGSSYKESSEFTVPGTKIVS--VDSVPGRLGILT 186

237 VCYDMRPELELALA-QNGAEILTYPSAFSGSITGPAHWEVLLRAITGCVVAAACG 295

187 VCYDLRFPKIYQQLRFEGKAOVLVPSAFVTVTGEAHWEILLRAVLETCQVIAAQAQ 246

296 RHHEKRASGSHVNDPMGTAVARCS--GPGICLARIIDLNTYLRRLHLP 345

247 KHNEKREYSGTLTIIDPMGTAVGRPLDRVSTGIYVADIDFSLIDSVRKMP 298

#### RESULT 3

40601 putative nitrilase homolog - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C:Accession: T40601

R:Wood, V.; Rajandream, M.A.; Bartell, B.G.; Voickaert, G.

submitted to the EMBL Data Library, February 1999

A:Reference number: Z21940

A:Accession: T40601

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-276 <WOO>

A:Cross-references: EMBL:AL035570; PIDN:CAB37598.1; GSPDB:GN00067; SPDB:SPBC651.02

A:Experimental source: strain 972h-; cosmid c651

C:Genetics:

A:Gene: SPDB:SPBC651.02

A:Map position: 2

A:Introns: 49/3; 81/3; 226/3; 232/1

C:Superfamily: hypothetical protein YLR351c

#### Query Match

30.0%; Score 575.5; DB 2; Length 276;

Best Local Similarity 44.2%; Pred. No. 1e-42;

Matches 126; Conservative 50; Mismatches 86; Indels 23; Gaps 8;

81 LPLVAVCVSTPDKOONFKCALVREARLGAFLPEAFDPIARDPAETLHLE-SEP 139

1 MTLAAVADLNSGSLTKLNAICKEILISQAAGKACIFPEASDPTAINSDAILELTNHP 60

140 LGGKLEETOLARECGMLSLGFERGODWEQOTOKIYNCVILLNSK-----GAVVA 192

61 DCSKFIKRVRSATKHSIFVNI-CVHE-----PSKYN--KLNLNSLFIETPLHGETIS 110

193 TYRTHLCLDVEIPGQPMCESNSTMPGSPLESVPSTPAKIGLAICYMRPELELALA 252

Db 111 RYKNAHLEDEVIKNGPPLKESNTLRCGAILLPCKTPIGKVGSAICFDIRPEQAIIKRN 170

253 AGAEILTYPSAFSGSITGPAHWEVLLRAVLETCQVVAACGRRHHEKRASGSHVNDP 312

171 MGAHITTPSAFTEKTGAHWEVLLRAVLETCQVVAACGRRHHEKRASGSHVNDP 230

313 WGVTVARCS--GP-GICLARIIDLNTYLRRLHLPVQHRPPDY 354

231 WGVTVAGYSDISPSNGLIFADLDNLVDHVRITYPLT--RRNDLY 273

#### RESULT 4

B69109

N-carbamoyl-D-amino acid amidohydrolase - *Methanobacterium thermoautotrophicum* (strat

C:Species: *Methanobacterium thermoautotrophicum*

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999

C:Accession: B69109

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.

; Qiu, D.; Spadafora, R.; Vitale, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivanli,

k1, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: fu

A:Reference number: A69000; MUID:98037514

A:Accession: B69109

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-272 <MTH>

A:Cross-references: GB:AE000934; GB:AE000666; NID:32622924; PIDN:AB86277.1; PID:3262

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1811

C:Superfamily: hypothetical protein YLR351c

#### Query Match

28.0%; Score 537.5; DB 2; Length 272;

Best Local Similarity 38.4%; Pred. No. 2.3e-39;

Matches 109; Conservative 49; Mismatches 95; Indels 31; Gaps 4;

84 VAVCVSTPDKOONFKCALVREARLGAFLPEAFDPIARDPAETLHLEPLGK 143

3 LGICOMVTEDEKERNKVGAVEMIVAAERGARLYLPEMF-----TCPIYVN 49

144 LLEEYTO-----LARECGMLSLGFERGODWEQOTOKIYNCVILLNSKGA 191

50 LFEYAEDENGYSIRTMRSIARELGIHLVAGSIDE-----ETPEGIYNTSVMDGNNI 104

192 ATYRTHCLDVEIPGQPMCESNSTMPGSPLESVPSTPAKIGLAICYMRPELELALA 251

105 GKRRVHLFDINVPGLSFRSDSLAGDSV--TVETQCYVAGVICTDMRFPPLSRMA 163

252 QNGAEILTYPSAFSGSITGPAHWEVLLRAVLETCQVVAACGRRHHEKRASGSHVND 311

164 LGGAELVLFPGAFNMNTTIPAHMRLVRSRALDNCYCAVAPARNPSASYAAGSHMAD 223

312 PWGTVVARCSGPGICLARIIDLNTYLRRLHLPVQHRPPDY 355

224 PWGTVCEADASPVTADIDLEAVERIRRLPLRLRRRPDVG 267

#### RESULT 5

S56907

hypothetical protein YUL126w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein J0706

C:Species: *Saccharomyces cerevisiae*

C>Date: 05-May-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Sep-1999

C:Accession: S56907

R:Czepluch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.

submitted to the Protein Sequence Database, September 1995

A:Accession: S56891

A:Molecule type: DNA

A:Residues: 1-307 <CZ1>

A:Cross-references: EMBL:Z49401; NID:g1008323; PIDN:CAA89421.1; PID:g1008324; MIPS:YU







A; Experimental source: strain A3(2)  
C; Genetics:  
A; Gene: SCOEDB:SCGD3.08c  
C; Superfamily: hypothetical protein YLR351c

Query Match	17.8%	Score 342.5	DB 2	Length 264
Best Local Similarity	35.8%	Pred. No. 2.2e-22		
Matches 91	Conservative 33	Mismatches 101	Indels 29	Gaps 9

QY	103	AEVLEEARLACACLAAPRE-----APFIRADPAETHLESEPGKLEBYTOLARECG	157
Db	24	AALVREGA--GADLVLEBELMTTGAFAFEEERDA-----AEPLRGPIYEMAMAAADAGV	76
QY	158	WLSLGGFHEHGQDMQFOKLYNCHVLLNSKGAVALYATRYKTHLCDVETIPGQPMCESNSTM	217
Db	77	WLHAGSVBERPD-----GRUTNSLIVSPAGDULASTYRKTHREGFD-----KGEAVL	124
QY	218	PGPSPLESVSVR--PAGKIGLAVCYDMFPEPLSTALAAQAGAILTYPARSGSITGPAHWEV	275
Db	125	MGAGRE--PYVTRLEPDTTLGVAVTCIDLEAFPEPLSRSLVDAGAEILLVPPAGWE--RRRHWTLL	182
QY	276	LJLRARAIETQCYVVAALAOCCGHHNHKRRASYSCHSNVDPWGTGVVVARCSEGPICLARIIDLNY	335
Db	183	LARARAVENQAFVLACGTAAGTHAGVPA--GHSIYVDPWGEVLAVAGADEVYLAVERDPER	241
QY	336	LRLRLRLPLRVQFNR	349
Db	242	VARTREDFPALKDR	255

RESULT 15  
F75263  
Probable hydrolase - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: F75263  
R:White, O.; Eelsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.: Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma  
S.: Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: F75263  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-297 <WHIT>  
A:Cross-references: GB:AE002082; GB:AE000513; NID:96460347; PIDN:AAF12070.1; PID:96460351  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2535  
A:Map position: 1  
C:Superfamily: hypothetical protein YLR351c

Query Match	16.6%	Score 318	DB 2	Length 297
Best Local Similarity	31.3%	Pred. No. 3.5e-20		
Matches	92	Conservative 49	Mismatches 117	Indels 36
			Gaps	10

  

QY	84	VAVCVYSTPRDKQNFKCAELVREARLGLACLAFLPEAFD--FLARDPAETHLSEPL	140
		: : : : :     : : : : :     : : : : :     : : : : :	
Db	10	LAVVQMHTDLEDVNERAAEHVREARAGAVILLPELFENLYFCOVEREDYFGSLAHL	69
		: : : : :     : : : : :     : : : : :     : : : : :	
QY	141	-GKLLLEVYQTLARECGMLSLGCGHNERGODEOTOKIYNCVLLNSKAVVAATYRKTL	199
		: : : : :     : : : : :     : : : : :     : : : : :	
Db	70	EGHPIGRFQELARELNVLLPVSYFEKAG-----QAHNLSLVCIDAGSELLGNRKTHI	123
		: : : : :     : : : : :     : : : : :     : : : : :	
QY	200	CDVEIPGGGPMQESNSTM-PPGSLSPEVSTPAKRGICLVCMYMRPELSLAQAQAEFL	258
		: : : : :     : : : : :     : : : : :     : : : : :	
Db	124	PD-----GPGVEEYKIFRPPDGTGKRWMDTRGRGAVGICMOMIPETARVAMLOGADFL	177
		: : : : :     : : : : :     : : : : :     : : : : :	
QY	259	TPPSAFSGITGPAH-----WEVLLRARAIETQCVVAVAAGGRH--HEKRASYGH	306
		: : : : :     : : : : :     : : : : :     : : : : :	

Db 178 LYTAIGS - EPAVEYEPNNIQQMRAMVGHAVSNSSYVGSSNNIGKEIYGGLEQTYIGH 235

Qy 307 SMMYDPKCTVYAR --- CSEGRGLCLARTIDLYLQRLRHLPVFOHRRRPDLXGNL 357

Db 236 SFISDYGELVALAELGDSDEGR - LIHLENLKEARKFRAGMGFFPRDRPELYGRL 287

Search completed: April 26, 2002, 16:55:52  
Job time: 180 sec





THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:54:36 ; Search time 56.94 Seconds

(without alignments)  
470.926 Million cell updates/sec

Title: US-09-357-675C-25

Perfect score: 1921

Sequence: 1 PLAAAXLAPRPPDRLRMV.....LPVFQHRPDLVGNLGHPLS 362

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:\*  
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1919	99.9	464	21	AAV68739
2	922	48.0	224	21	AAV687054
3	821	42.7	153	22	AAV10877
4	485.5	25.3	276	22	AAV80984
5	310	16.1	159	21	AAV42967
6	283.5	14.8	266	22	AAV92814
7	270	14.1	299	21	AAV09683
8	256.5	13.4	299	21	AAV4945
9	256.5	13.4	252	22	AAV79700
10	256.5	13.4	252	22	AAV80193
11	241	12.5	220	21	AAV09684

12	241	12.5	220	21	AAV49946	Arabidopsis thalia
13	218	11.3	264	22	AAV81841	S. epidemidis ope
14	217.5	11.3	340	21	AAV44266	Arabidopsis thalia
15	217.5	11.3	346	21	AAV44265	Arabidopsis thalia
16	217.5	11.3	358	21	AAV44264	Arabidopsis thalia
17	199.5	10.4	303	15	AAV46260	Improved Heat-stab
18	199.5	10.4	303	15	AAV46258	Improved Heat-stab
19	199.5	10.4	303	15	AAV46251	Heat-stable carbam
20	199.5	10.4	352	21	AAV05762	Arabidopsis thalia
21	198.5	10.3	303	15	AAV46241	Heat-stable carbam
22	198.5	10.3	303	15	AAV46268	Improved Heat-stab
23	198.5	10.3	339	21	AAV05763	Arabidopsis thalia
24	197.5	10.3	303	15	AAV46269	Improved Heat-stab
25	197.5	10.3	303	15	AAV46270	Improved Heat-stab
26	196.5	10.2	303	15	AAV46249	Heat-stable carbam
27	196.5	10.2	303	15	AAV46257	Improved Heat-stab
28	196.5	10.2	303	15	AAV46255	Improved Heat-stab
29	196.5	10.2	303	15	AAV46261	Improved Heat-stab
30	196.5	10.2	303	15	AAV46259	Improved Heat-stab
31	196.5	10.2	303	15	AAV46243	Heat-stable carbam
32	196.5	10.2	303	15	AAV46244	Heat-stable carbam
33	196.5	10.2	303	15	AAV46265	Improved Heat-stab
34	196.5	10.2	303	15	AAV46267	Improved Heat-stab
35	196.5	10.2	303	15	AAV46272	Improved Heat-stab
36	196.5	10.2	303	22	AAV74625	Escherichia coli d
37	195.5	10.2	303	15	AAV46247	Heat-stable carbam
38	195.5	10.2	303	15	AAV46248	Heat-stable carbam
39	195.5	10.2	303	15	AAV46252	Heat-stable carbam
40	195.5	10.2	303	15	AAV46253	Heat-stable carbam
41	195.5	10.2	303	15	AAV46250	Improved Heat-stab
42	195.5	10.2	303	15	AAV46262	Improved Heat-stab
43	195.5	10.2	303	15	AAV46263	Heat-stable carbam
44	195.5	10.2	303	15	AAV46239	Heat-stable carbam
45	195.5	10.2	303	15	AAV46240	Heat-stable carbam

#### ALIGNMENTS

RESULT 1	
ID	AAV68739 standard; Protein; 464 AA.
XX	
AC	AAV68739;
XX	
DT	05-MAY-2000 (first entry)
XX	
DE	Amino acid sequence of the N11 protein.
XX	
KW	N11 gene; nitrlase; tumour suppressor gene; FHIT; chromosome 3p14.2;
KW	FRA3B; cancer; genome allele inactivation.
XX	
OS	Homo sapiens.
OS	Mus sp.
OS	Drosophila melanogaster.
OS	Caenorhabditis elegans.
XX	
FT	Key
FT	Misc-difference
FT	Location/Qualifiers
FT	6 /note= "unspecified amino acid encoded by TNT"
PN	WO200003685-A2.
XX	
PD	27-JAN-2000.
XX	
PF	20-JUL-1999; 99WO-US16366.
XX	
PR	20-JUL-1998; 98US-0093350.
XX	
PA	(UYJE-) UNTV JEFFERSON THOMAS.
XX	
PI	Croce CM;
XX	

DR WP1: 2000-171195/15.  
DR N-PSDB: AAZ46101.  
XX  
PT Novel nitrlase homologs used as diagnostic and therapeutic reagents  
PT for the detection and treatment of cancer -  
XX  
PS Disclosure: Fig 6; 25pp; English.  
XX  
CC The present sequence is encoded by the coding region of human, murine,  
CC Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The  
CC human and mouse NIT1 genes are members of an uncharacterised  
CC mammalian gene family with homology to bacterial and plant nitrlases.  
CC The tumour suppressor gene FHIT in D. melanogaster and C. elegans code  
CC for fusion proteins in which the Phit domain is fused with a Nit domain.  
CC In mouse and humans, FHIT and NIT are encoded by two different genes,  
CC localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The  
CC human FHIT gene at chromosome 3p14.2, spanning the constitutive  
CC chromosomal fragile site FRA3B, is often altered in most common forms  
CC of human cancer. The NIT1 protein overcomes the mutated inactivation  
CC of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives  
CC and analogues of them, and antibodies are used as diagnostic and  
CC therapeutic reagents for the detection and treatment of cancers.  
XX  
SQ Sequence 464 AA:  
  
Query Match 99.9%; Score 1919; DB 21; Length 464;  
Best Local Similarity 100.0%; Pred. No. 3,1e-189;  
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PLAAAXLAPDRPPDRTLRLMVLAISSCRYSLSRRRLGFTTRPPHRLSLCPGLRIPL 60  
DB 1 plaaaxlapdrppdrtllrmvlaisscryslarrprlglftrpphrfslslcpglrlpql 60  
QY 61 SVLCAQPPRRMAISSSCCEPLVAVCOVTSPPDKQONFKTAELVREARAGACLAFLP 120  
DB 61 svlcqpprrmaisssscelpvavcvtsppdkqnfktcaelvreaarlagacflap 120  
QY 121 EAFDRIADPAETLHLSPILGKLEEYQTOLARECGMLWSLGFHREGODMEQOTKIYNC 180  
DB 121 eafdiardpaetlhlspilgkleeeyqtolarecgmlwslygfhergdmeqotkiyinc 180  
QY 121 eafdiardpaetlhlspilgkleeeyqtolarecgmlwslygfhergdmeqotkiyinc 180  
DB 121 eafdiardpaetlhlspilgkleeeyqtolarecgmlwslygfhergdmeqotkiyinc 180  
QY 181 HVLLNSKGAVVATYRKTHLCDEVEIPGQPMCESNSTMGPSPLESFVSTPAGKIGLAVCYD 240  
DB 181 hvllnsgavvatyrkthlcdeveipgqpmcesnstmgpsslesfvstpagkiglavcyd 240  
QY 241 MRFPELSIALAQAGAEITLTPSAFSGITGPAHWEVLRAARAEITOCYVAAACGRHNEK 300  
DB 241 mrfpelsialaqagaeiltypsafigitgpahevllraaraeitletcyvaacqgrhnek 300  
QY 301 RASYGHSWVDPWGTGVAVARCSGPGICLARIDLNLRLRHLPVFOHRRPDLVGNLGHF 360  
DB 301 rasylghswvdpwgtgvavarcsegpgiclaridlnlrlrhlpvfohrrpdlvgnlghf 360  
QY 361 LS 362  
DB 361 ls 362  
  
RESULT 2  
AAB57054  
ID AAB57054 standard; Protein: 224 AA.  
XX  
AC AAB57054;  
XX  
DT 13-MAR-2001 (first entry)  
XX  
DE Human prostate cancer antigen protein sequence SEQ ID NO:1632.  
XX  
KW Human: prostate cancer; prostate cancer antigen: detection; diagnosis;  
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;  
KW vulnereary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;

KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease.  
XX  
OS Homo sapiens.  
XX  
PN WO20005174-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05988.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
PI Rosen CA, Ruben SM;  
XX  
DR WP1: 2000-587513/55.  
DR N-PSDB: AAF16257.  
XX  
PT Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -  
XX  
PS Claim 11; Page 2090-2091; 2338pp; English.  
XX  
CC AAF15566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB5363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytosolic,  
CC cardioactive, immunomodulatory, muscular, vulnereary, gastrointestinal,  
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.  
XX  
SQ Sequence 224 AA:  
  
Query Match 48.0%; Score 922; DB 21; Length 224;  
Best Local Similarity 94.7%; Pred. No. 8e-87;  
Matches 178; Conservative 3; Mismatches 3; Indels 4; Gaps 3;  
  
QY 37 LGFTRPPHRLSLCPGLRIPLSVLCAOPRRMAISSSCCEPLVAVCOVTSPPDKQ 96  
DB 27 lgftrpphrfslslcpglrlpqlsvlcavqprpmaisssscelpvavcvtsppdkq 96  
QY 97 ONFKTCAELVREARAGACLAFLPAEPDFIARDPAETLHLSPILGKLEEYQTOLARECG 156  
DB 87 onfktaelvreaarlagacflpaeidfiaardpaetlhlspilgkleeeyqtolarecg 146  
QY 157 LWSLGFHREGODMEQOTKIYNCVLLNSKGAVVATYRKTHLCDEVEIPGQPMCESNST 216  
DB 147 lwslygfhergdmeqotkiyincvllnsgavvatyrkthlcdeveipgq-ic-vvat 204  
QY 217 MP--GPSTL 222  
DB 205 lpcslpvl 212  
  
RESULT 3  
AAI10877  
ID AAI10877 standard; Protein: 153 AA.  
XX  
AC AAI10877;  
XX  
DT 14-MAY-1999 (first entry)

XX Amino acid sequence of a human secreted protein.  
 DE  
 XX  
 KW Secreted protein; cancer; tumour; neurodegenerative disorder;  
 KW developmental abnormality; foetal deficiency; blood disorder;  
 KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;  
 KW renal disease; diabetes; inflammation; allergy; ischemic shock;  
 KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;  
 KW prostate disease; asthma; osteoporosis; arthritis.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..46  
 FT /note= "signal peptide"  
 FT Protein 47..153  
 FT /note= "secreted protein"  
 XX  
 PN W09907891-A1.  
 PD 18-FEB-1999.  
 PF 04-AUG-1998; 98WO-US16235.  
 XX  
 PR 19-AUG-1997; 97US-0056732.  
 PR 05-AUG-1997; 97US-0054798.  
 PR 05-AUG-1997; 97US-0054803.  
 PR 05-AUG-1997; 97US-0054804.  
 PR 05-AUG-1997; 97US-0054806.  
 PR 05-AUG-1997; 97US-0054807.  
 PR 05-AUG-1997; 97US-0054808.  
 PR 05-AUG-1997; 97US-0054809.  
 PR 05-AUG-1997; 97US-0055309.  
 PR 05-AUG-1997; 97US-0055310.  
 PR 05-AUG-1997; 97US-0055312.  
 PR 05-AUG-1997; 97US-0055386.  
 PR 05-AUG-1997; 97US-0055311.  
 PR 18-AUG-1997; 97US-0055970.  
 PR 18-AUG-1997; 97US-0055986.  
 PR 19-AUG-1997; 97US-0056365.  
 PR 19-AUG-1997; 97US-0056366.  
 PR 19-AUG-1997; 97US-0056557.  
 PR 19-AUG-1997; 97US-0056370.  
 PR 19-AUG-1997; 97US-0056371.  
 PR 19-AUG-1997; 97US-0056563.  
 PR 19-AUG-1997; 97US-0056731.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Brewer LA, Ehner R, Ferrie AM, Greene JM, Janat F, N1 J;  
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;  
 DR WPI; 1999-167452/14.  
 DR N-PSDB; AAX30398.  
 XX  
 PT New isolated human genes encoding secreted polypeptides - useful for  
 PT diagnosis and treatment of pathological diseases  
 XX  
 PS Claim 3; Page 304; 331pp; English.  
 XX  
 CC The specification describes secreted proteins and their corresponding  
 CC polynucleotides which are useful for preventing, treating or ameliorating  
 CC medical conditions, e.g. by protein or gene therapy. Pathological  
 CC conditions can also be diagnosed by determining the amount of the  
 CC secreted polypeptides in a sample or by determining the presence of the  
 CC mutations in the polynucleotides. Specific uses are described for each  
 CC of the products, based on which tissues they are most highly  
 CC expressed in, and include developing products for the diagnosis or  
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental  
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,  
 CC diseases of the immune system, autoimmune diseases, hepatic and renal  
 CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's  
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,

CC prostate diseases, asthma, disorders involving osteoclasts such as  
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or  
 CC thymus, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 XX  
 SQ Sequence 153 AA;

Query Match 42.7%; Score 821; DB 20; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-76;  
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 MCESNSTMPGSLSPSTPAGKIGLAVCYDMFPELTLAOGAEILTPSAFGSITG 269  
 Db 1 mcesnstmpgslspstpapgkiglavcydmfpelstlaogaeiltpsafigsitg 60  
 QY 270 PAHWEVLLRRARLETQCYVAAACGRHBEKRASYSCHSWVDPWGVYVARCSGGGLCLA 329  
 Db 61 pahwevllrrarletqcyvaaaqgrhbkcrasyshsmvdpwgtvarcsegpglcla 120  
 QY 330 RIDLNTYRQLRRHLVPVQHRPDLXGTLGHPLS 362  
 Db 121 ridlntyrqlrrhlvpvqhrpdlxgtnlghpls 153

## RESULT 4

AAB80984  
 ID AAB80984 standard; Protein; 276 AA.

XX  
 AC AAB80984;

XX  
 DT 12-JUN-2001 (first entry)

XX  
 DE Human hnit3-ase.

XX  
 KW Human: hnit3-ase; Cushings adrenal gland tumour;

KM nitrile hydrolytic enzyme; nitrilase.

XX  
 OS Homo sapiens.

XX  
 PN CN1277998-A.

XX  
 PD 27-DEC-2000.

XX  
 PF 30-MAY-2000; 2000CN-0116221.

XX  
 PR 30-MAY-2000; 2000CN-0116221.

XX  
 PA (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.

XX  
 PI Xu X, Qian B, Zhang X;

XX  
 DR WPI; 2001-245678/26.

DR N-PSDB; AAF84214.

XX  
 PT Human nitrile hydrolytic enzyme protein and its coding sequence -

XX  
 PS Claim 2; Page 11; 20pp; Chinese.

XX  
 CC The present invention relates to human hnit3-ase (nitrilase) protein,  
 CC which is expressed in Cushings adrenal gland tumour, and its coding  
 CC sequence (AAF84214 and AAB80984). The present invention also relates to a  
 CC preparation method of said protein and nucleic acid sequence, and a  
 CC method of detecting human hnit3-ase nucleic acid sequence and polypeptide  
 CC in sample.  
 XX  
 SQ Sequence 276 AA;

Query Match 25.3%; Score 485.5; DB 22; Length 276;  
 Best Local Similarity 36.7%; Pred. No. 1.2e-41;  
 Matches 105; Conservative 46; Mismatches 98; Indels 37; Gaps 6;



XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 6568; 246pp + Sequence Listing; English.

XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 266 AA:

Query Match 14.8%; Score 283.5; DB 22; Length 266;

Best Local Similarity 32.3%; Pred.No. 7.9e-21; Mismatches 115; Indels 33; Gaps 10;

Matches 90; Conservative 41; Mismatches 115; Indels 33; Gaps 10;

QY 84 VAVQVSTPDKQONFKTCALREARACIAPFEAFDFARPATLHL---SEPL140  
DB 3 iallqtsndkmdnfalldaekaeqgarlvfpea---tsqsfygrldtgaeei 58

QY 141 GCKLEETVQLARECGMLSLGCFH----ERGDMEQOTOKITNCHVLSKCAVATYR 195  
DB 59 dgefstavkladelvlnvvaigmfpadvgrg---ektlsrvnmvlvsag-lhgyyn 114

QY 196 KTHLCVEIIPGQPMCESTMPGSPLESFVSPACKIGIACVYDMPFELSLALAQGA 255  
DB 115 khhkyda-----fgyresdtvxpqdel-vvfevddlkfgyatcydlrfpeqfxdlarnga 168

QY 256 EILTVPSAFGSITGPA---HMEVLLRARALETQCYVVAACGSRHNEKR-----ASYGH 306  
DB 169 qllvtpslsqgd--gpgklqgwevlpraraldscwlvacgqarlpreelrderkpgpigh 226

QY 307 SNVDPFWGVVAVARCEGPGICLARIDLNTYRQIRHLPLV 345  
DB 227 smvtnphgevlasagyepemlldidvsgjaklrealpv 265

RESULT 7

AAAG09683  
ID AAG09683 standard; Protein; 299 AA.

AC AAG09683;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 7710.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142054.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148177.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151388.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158368.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.1%; Score 270; DB 21; Length 299;  
Best Local Similarity 27.6%; Pred. No. 2,3e-19;  
Matches 85; Conservative 47; Mismatches 124; Indels 52; Gaps 10;

QY 83 LVAVCVSTPDKQONFRTGACELVREARLGCALFLPEAF-----DFIARPPA 131  
DB 11 vvsliqfacsdsltnvaaaerlvreaahkaganilllqelfeggyfcqagredffkrakp 70  
QY 132 ETLHLSEPLGKLEBYTQALARECGIMLSLGFHERGQDWEQTOKIYNCHVLLNSGAVV 191  
DB 71 yknh-----ptlarmqklakelgvipvsffe-----eantahynsialidagtdl 117  
QY 192 ATYRKTILCDVEIRPGQPRCESENSTM-PPPSLESFVSVPACRTGLAVCYDMKFFELSLAL 250  
DB 118 gilyrkshipd-----gpdygekfyfnpqdtgfkvfqtkfaklgaivalcwdqwfpeaarfam 171  
QY 251 AOGAEILTYPSAFGS-----ITGPAHWEVILRARAIETQCYVVAQAQGR-----HH 298  
DB 172 vlgaeellfyptaisgpeqdgldsrdrhrrymqghaganvpylvasnrlgkellletelv 231  
QY 299 EKRRAS-YGHSWVVDPMGVVAVRCSE-GPGLCLARIIDLNLQRLRHLPLVFOHRRPDLV-- 354  
DB 232 psgltfygtsflagptglvtaeaddkseeavlvaqfdldmikckrsgwgyfdrtrpdllykv 291  
QY 355 -----GNL 357  
DB 292 lltmgnl 299

RESULT 8  
AAG4945  
ID AAG4945 standard; Protein: 299 AA.  
XX AAG4945;  
AC  
XX  
XX 18-OCT-2000 (first entry)  
DT  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63238.  
DE  
XX





```
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 14.18; Score 270; DB 21; Length 299;  
Best Local Similarity 27.68; Pred. No. 2.3e-19;  
Matches 85; Conservative 47; Mismatches 124; Indels 52; Gaps 10;

```
OY 83 LVAVQVSTPDKQONFKVCAELVREARLACLAFLPEAF-----DFTARDPA 131
DB 11 VVSLQFACSDISTCNVAAEALIVEAHAKGANILLIGELFEGYYFCGQREDFFKRAKP 70
OY 132 ETLHLEPGLGRLBEYQOLARECGIMLSIGFHERGQDMEOTOKIYNCHVILNSGAVV 191
DB 71 YKbn-----ptlarmqklakelgvnvpvstfe-----eantahynsialiddgdcl 117
OY 152 ATYRTHLCDVEIRPGQCPWCSNSTM-FQPSLESVSTPAGTIGLAVCYDMFPPELISLAL 250
DB 118 glyrkslhpD-----gpygqekfyfnpdytdgfkvfqekfaklgvalicwdqyfpcaaram 171
OY 251 AQAQAEILTYPSAFGS-----ITGPAHMEVLLRARALFTQCVVVAQAQGR-----HH 298
DB 172 vlgaeellfypalaigsepqdgldsrldwrrvmqghaganvvpilvasnllkxellcteng 231
OY 299 EKRAAS-YGHSNVDWPGTVYVARCSE-GPGCLARIDLVNROLRHLRPFQHRPPDLX-- 354
DB 232 psqitfygtsflagptglvetaeaddkseavlvagfdlmlkxkqsgwyvfrdrpdllykv 291
OY 355 -----GNL 357
DB 355 -----GNL 357
```

```
DB 292 llumdn1 299
RESULT 9
ID AAB79700 standard; Protein; 252 AA.
AC AAB79700;
XX 30-APR-2001 (first entry)
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:134.
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme.
XX
OS Corynebacterium glutamicum.
PN
XX
XX WO200100843-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-IB00923.
XX
XX 25-JUN-1999; 99US-0141031.
XX 01-JUL-1999; 99DE-1030476.
XX 02-JUL-1999; 99US-0142101.
XX 08-JUL-1999; 99DE-1031415.
XX 08-JUL-1999; 99DE-1031418.
XX 08-JUL-1999; 99DE-1031419.
XX 08-JUL-1999; 99DE-1031420.
XX 08-JUL-1999; 99DE-1031424.
XX 08-JUL-1999; 99DE-1031428.
XX 08-JUL-1999; 99DE-1031434.
XX 08-JUL-1999; 99DE-1031435.
XX 08-JUL-1999; 99DE-1031443.
XX 08-JUL-1999; 99DE-1031453.
XX 08-JUL-1999; 99DE-1031457.
XX 08-JUL-1999; 99DE-1031465.
XX 08-JUL-1999; 99DE-1031478.
XX 08-JUL-1999; 99DE-1031510.
XX 08-JUL-1999; 99DE-1031541.
XX 08-JUL-1999; 99DE-1031573.
XX 08-JUL-1999; 99DE-1031592.
XX 08-JUL-1999; 99DE-1031632.
XX 08-JUL-1999; 99DE-1031634.
XX 08-JUL-1999; 99DE-1031636.
XX 09-JUL-1999; 99DE-1032125.
XX 09-JUL-1999; 99DE-1032126.
XX 09-JUL-1999; 99DE-1032130.
XX 09-JUL-1999; 99DE-1032186.
XX 09-JUL-1999; 99DE-1032206.
XX 09-JUL-1999; 99DE-1032227.
XX 09-JUL-1999; 99DE-1032228.
XX 09-JUL-1999; 99DE-1032229.
XX 09-JUL-1999; 99DE-1032230.
XX 14-JUL-1999; 99DE-1032922.
XX 14-JUL-1999; 99DE-1032926.
XX 14-JUL-1999; 99DE-1032928.
XX 14-JUL-1999; 99DE-1033004.
XX 14-JUL-1999; 99DE-1033005.
XX 14-JUL-1999; 99DE-1033006.
XX 12-AUG-1999; 99US-0148613.
XX 27-AUG-1999; 99DE-1040764.
XX 27-AUG-1999; 99DE-1040765.
XX 27-AUG-1999; 99DE-1040766.
XX 27-AUG-1999; 99DE-1040832.
XX 31-AUG-1999; 99DE-1041378.
XX 31-AUG-1999; 99DE-1041379.
XX 31-AUG-1999; 99DE-1041380.
```

```
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
XX (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-137957/14.
DR N-PSDB; AAF71819.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
XX pathway proteins, useful for producing fine chemicals in
XX microorganisms, including organic acids, nonproteinogenic amino acids,
XX and purine and pyrimidine bases -
XX
XX Claim 20; Page 361-362; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
XX pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
XX MP nucleic acids are useful for the production of fine chemicals
XX in microorganisms, including organic acids, nonproteinogenic amino
XX acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
XX saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
XX compounds, vitamins, cofactors, polyketides and enzymes.
XX
XX Sequence 252 AA;
SQ
Query Match 13.4%; Score 256.5; DB 22; Length 252;
Best Local Similarity 32.5%; Pred. No. 4,5e-18;
Matches 86; Conservative 36; Mismatches 110; Indels 33; Gaps 10;
QY 98 NFKTCAELVREARAGACIAFLPEAFDFIARDBAEFLHL---SEPLGKLEERYTDLARE 154
DB 3 nfallrrdaaeakaegarylvfpea---tsqsfgyridtaeeeldegfstavrxklade 58
QY 155 CGIMLSLIGGFH-----ERGDWEQOTOKIYNCHVLNSKGAVVATYRKTHLCDEIRGQGP 209
DB 59 ldvniivaqmftrpadctvgir---ektstynntvllisgag-llnggykthtyda-----fg 109
QY 210 KCESNSTMPGSPLESFVSTPAGKIGIAVCYDMKFPPLSLAOLAQGAELITTPSAFGSITG 269
DB 110 yresdtkvpgdel-vvfevddikfyatcgydtrfpeqfkdlarngsqllvprswqd--g 166
QY 270 PA---HWEVLTLARALETQCYVAAAGCGRHNEKR-----ASYGSMVVDPMGTIVANC 320
DB 167 psklqewelpirataedscwliwacqgarlpeeliderkprtyglghsmvtnprgveiaaa 226
QY 321 SEGPGICLARIDNLNLRRLPLPV 345
DB 227 gyeemlladidvsglakitrealpr 251
RESULT 10
AAB80193
ID AAB80193 standard; Protein; 252 AA.
XX
XX AAB80193;
XX
XX 30-APR-2001 (first entry)
XX
XX Corynebacterium glutamicum MP protein sequence SEQ ID NO:1120.
XX
```

```
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX
XX Corynebacterium glutamicum.
XX WO200100843-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-IB00923.
XX
XX 25-JUN-1999; 99US-0141031.
XX 01-JUL-1999; 99DE-1030476.
XX 02-JUL-1999; 99US-0142101.
XX 08-JUL-1999; 99DE-1031415.
XX 08-JUL-1999; 99DE-1031418.
XX 08-JUL-1999; 99DE-1031419.
XX 08-JUL-1999; 99DE-1031420.
XX 08-JUL-1999; 99DE-1031424.
XX 08-JUL-1999; 99DE-1031428.
XX 08-JUL-1999; 99DE-1031434.
XX 08-JUL-1999; 99DE-1031435.
XX 08-JUL-1999; 99DE-1031443.
XX 08-JUL-1999; 99DE-1031453.
XX 08-JUL-1999; 99DE-1031457.
XX 08-JUL-1999; 99DE-1031465.
XX 08-JUL-1999; 99DE-1031478.
XX 08-JUL-1999; 99DE-1031510.
XX 08-JUL-1999; 99DE-1031541.
XX 08-JUL-1999; 99DE-1031573.
XX 08-JUL-1999; 99DE-1031592.
XX 08-JUL-1999; 99DE-1031632.
XX 08-JUL-1999; 99DE-1031634.
XX 08-JUL-1999; 99DE-1031636.
XX 09-JUL-1999; 99DE-1032125.
XX 09-JUL-1999; 99DE-1032126.
XX 09-JUL-1999; 99DE-1032130.
XX 09-JUL-1999; 99DE-1032186.
XX 09-JUL-1999; 99DE-1032206.
XX 09-JUL-1999; 99DE-1032227.
XX 09-JUL-1999; 99DE-1032228.
XX 09-JUL-1999; 99DE-1032229.
XX 09-JUL-1999; 99DE-1032230.
XX 14-JUL-1999; 99DE-1032922.
XX 14-JUL-1999; 99DE-1032926.
XX 14-JUL-1999; 99DE-1033028.
XX 14-JUL-1999; 99DE-1033004.
XX 14-JUL-1999; 99DE-1033005.
XX 14-JUL-1999; 99DE-1033006.
XX 12-AUG-1999; 99US-0148613.
XX 27-AUG-1999; 99DE-1040764.
XX 27-AUG-1999; 99DE-1040765.
XX 27-AUG-1999; 99DE-1040766.
XX 27-AUG-1999; 99DE-1040832.
XX 31-AUG-1999; 99DE-1041378.
XX 31-AUG-1999; 99DE-1041379.
XX 31-AUG-1999; 99DE-1041380.
XX 31-AUG-1999; 99DE-1041394.
XX 31-AUG-1999; 99DE-1041396.
XX 03-SEP-1999; 99DE-1042076.
XX 03-SEP-1999; 99DE-1042077.
XX 03-SEP-1999; 99DE-1042079.
XX 03-SEP-1999; 99DE-1042086.
XX 03-SEP-1999; 99DE-1042087.
XX 03-SEP-1999; 99DE-1042088.
XX 03-SEP-1999; 99DE-1042095.
XX 03-SEP-1999; 99DE-1042124.
XX 03-SEP-1999; 99DE-1042129.
XX 09-MAR-2000; 2000US-0187970.
XX
```

PA (BADI ) BASE AG.  
XX  
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
XX WPI: 2001-137957/14.  
DR N-PSDB; AAF72312.  
XX  
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
PT pathway proteins, useful for producing fine chemicals in  
PT microorganisms, including organic acids, nonproteinogenic amino acids,  
PT and purine and pyrimidine bases -  
XX  
XX Claim 20; Page 1678-1679; 1737pp; English.  
PS  
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
CC MP nucleic acids are useful for the production of fine chemicals  
CC in microorganisms, including organic acids, nonproteinogenic amino  
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
CC compounds, vitamins, cofactors, polyketides and enzymes.  
XX  
XX Sequence 252 AA;  
SQ  
  
Query Match 13.4%; Score 256.5; DB 22; Length 252;  
Best Local Similarity 32.5%; Pred. No. 4.5e-18;  
Matches 86; Conservative 36; Mismatches 110; Indels 33; Gaps 10;  
  
QY 93 NFKTCAELVREARAGACLAFLPEAFDIARDPATLHL---SEPLGKLEFYQLARE 154  
Db 3 nfallrdaaeakagaavlvfpea---tsqsfyglrldtgaeldgfeslavvklade 58  
QY 155 GGLMLSLGCFH---ERGDWEOTOKIYNCHVLNSKGAVVATRKTHLDCVEIRPGGP 209  
Db 59 ldvvlvagamfipadvtvqg---ekistvntvlisgag-lhgyynkhltyda-----fg 109  
QY 210 MCESNSTMPGSPLESVPSTPAKIGLAVCYDMRPPELSIALAQAQAEILTYPSAFGSITG 269  
Db 110 yresdtvpgdel-vvfeavddikfgvatcgydlrfpeqfkdlarnagaqlivvptswgd--g 166  
QY 270 PA---HMEVLLRARAIEFQCIVVAAAGCGRHNEK-----ASYHSNWDVPWGVAVARC 320  
Db 167 pglkqewvvlpraraldstcwlacgqarlpaelrderkprtyghsmvtnphgevlasa 226  
QY 321 SEGPGICLARIDLNYLRRLRPV 345  
Db 227 gyepemliadivsglakirealpv 251  
  
RESULT 11  
ID AAG09684  
AAG09684 standard; Protein: 220 AA.  
XX  
XX AAG09684;  
AC  
XX  
XX 17-OCT-2000 (first entry)  
DT  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 7711.  
DE  
XX  
XX Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EPI033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR

PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.



[illegible]



DT 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 55427.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
PD  
XX 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137228.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.



```
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 11.3%; Score 217.5; DB 21; Length 340;  
Best Local Similarity 26.4%; Pred. No. 7,3e-14;  
Matches 87; Conservative 47; Mismatches 128; Indels 67; Gaps 14;

```
OY 76 SSSCEPLVAVQVSTPRKQNFKCALVREARLGAACLAFLPRAF----- 123
DB 16 sstvtvtlvgstvyv--dtpatlhaekfiyeaaekgaklviifpeafigyprgtrfgl 73
OY 124 -----DFIARDPAETLHSE--PLGSKLIEEYTOIARECGIMLSIGFFHFGQDWEQTOKI 177
DB 74 avgvhneegridelfrnyhaaakvprpxverlaelagknvhlvmgalekdg-----ylt 127
OY 178 YNCHVLLNSKGAHVATYRKTHLCDFE--IPGQGPMECSNSTMPGSLSEFVSTPAKIGL 235
DB 128 yctalffspqgflgkhrvtmptslercilwgq-----dgsstlp-----vydtpigkiga 177
OY 236 AYCQYMRPFELSLALQAQAELITVPSAFGSTTGRAHWEVLARARILEQCYYVAAQ-C 294
DB 178 alcwemrmljyrtalyakleiycaptadysl---ewqasmhlhvegcivlsahqfc 233
OY 295 GR-----HHEKRSYGHSMVVDPMGTVVARS--EGPCLCLARIDLNV 335
```

```
DB 234 krrfepohpylfndlvtckehqptvsggsvlispigkvlagpnyeseglvradldlga 293
OY 336 LROLRKRLPYFOH-RRPDLXGNL---GHP 360
DB 294 Iarakiyfdvghyskppdlf-nltvnehp 321

RESULT 15
AAC44265
ID AAC44265 standard; Protein; 346 AA.
XX
AC AAC44265;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55426.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
```

```
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147319.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149422.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0152370.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.3%; Score 217.5; DB 21; Length 346;
Best Local Similarity 26.4%; Pred. No. 7.5e-14;
Matches 87; Conservative 47; Mismatches 128; Indels 67; Gaps 14;

Qy 76 SSSCEPLVAVCGVSTPQKQNFKTCALVREARIGACIAFLPEAF----- 123
Db 22 sstvtvtivgsetvyn---dtpatlhkaekfiweaaskgakiylfpeafigyprgtrfgrl 79
Qy 124 ----DTIARDPAETLHUSE-PLGSKLLEETVYQTLARECGIWMISLGFFHGRGDWEQTORI 177
```

```

Db      80 avgvhneegridefrnyhasaalkvpqpxverlaelagknvhlvmgaiekdg-----ytl 133
QY     178 YNCHVLNSKGAVVATYRKTHLCVYE--IPGQGPWCESNSTMPGPSLESVPSTPAKIGL 235
Db     134 yctalfispqgqflgkhrkmpptslerciwggg----dgsitp-----vydtpigkiga 183
QY     236 AVCYDMRPPELSIALAAGAEILTYPSAFSGSITGPAHMEVILRARAIETQCYVVAAG-C 294
Db     184 aicwenrmplyrtalyakyleycaptadysl-----ewqasmihlaveggctvlisahqfc 239
QY     295 GR-----HHEKRASYGSHSNVDPWGTVVARCS-EGPGLCLARIDLNY 335
Db     240 krreipenhpylfndivdtkenhdpvsgsgsvlispjgkvlagpnyeseglvrtadidlgd 299
QY     336 LRQLRRHLVPFOH-RRPDLYGNL---GHP 360
Db     300 iaraklyfdvavghyskpdif-nltvnehp 327

```

Search completed: April 26, 2002, 16:54:37  
 Job time: 235 sec

**THIS PAGE BLANK (USPTO)**

OM of: US-09-357-675C-21 to: GenEmbl:\* out\_format : pfs

Date: Apr 29, 2002 11:05 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

#### Command line parameters:

-MODEL=frame+D2n.model -DEV=x1h  
-Q=/cgn1\_1/USPIO.spool/US09357675/rnat\_29042002\_091757\_3179/app-query.fasta.1.1208  
-DB=genEmbl -QMT=fastcap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs  
-NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09357675.@CGN1\_1.10065 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPRX -WAIT -THREADS=1

#### Search information block:

Query: US-09-357-675C-21  
Query length: 327

Database: GenEmbl:\*

Database sequences: 1797656

Search time (sec): 3374.960000

#### Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb.pr:AF069987	+ 1748.00	2415.67	3.4e-126	1385	AF069987 Homo sapiens nitrilas
gb.ro:AF069988	+ 1484.00	2049.81	8.2e-106	1388	AF069988 Mus musculus nitrilas
gb.ro:BC021634	+ 1476.00	2038.53	3.5e-105	1365	BC021634 Mus musculus nitrilas
gb.pr:AF069984	+ 1468.50	2018.45	4.6e-104	14079	AF069984 Homo sapiens nitrilas
gb.hlg:AL354714.4	+ 1468.50	1989.30	1.9e-102	110000	Continuation (5 of 7) of AL3
gb.hlg:AL354714.5	+ 1468.50	1989.30	1.9e-102	110000	Continuation (6 of 7) of AL3
gb.hlg:AL3590651	+ 1468.50	1985.56	3.1e-102	167863	AL3590651 Homo sapiens chrom
gb.pr:AL591806	+ 1468.50	1983.97	3.8e-102	200822	AL591806 Human DNA sequence
gb.hlg:AC084821	+ 1105.50	1480.02	4.5e-74	211771	AC084821 Mus musculus chrom
gb.ro:AF069985	+ 1102.50	1509.77	9.6e-76	4481	AF069985 Mus musculus nitrilas
gb.hlg:AC087229	+ 1083.00	1448.95	2.5e-72	213203	AC087229 Mus musculus chrom
gb.hlg:AC105589	+ 1059.00	1417.03	1.4e-70	178418	AC105589 Rattus norvegicus c
gb.ov:AF284575	+ 968.50	1335.82	4.9e-66	1214	AF284575 Xenopus laevis Nit p
gb.in:AF069989	+ 710.50	935.82	5.4e-46	1521	AF069989 Drosophila melanogast
gb.hlg:AC017755	+ 710.50	936.34	8.6e-44	131853	AC017755 Drosophila melanog
gb.in:AC093121	+ 710.50	934.18	1.1e-43	168250	AC093121 Drosophila melanog
gb.hlg:AC006169	+ 710.50	933.91	1.2e-43	173494	AC006169 Drosophila melanog
gb.in:AC105293	+ 710.50	933.09	1.3e-43	174279	AC105293 Drosophila melanog
gb.in:AE003467	+ 710.50	929.10	2.2e-43	190351	AE003467 Drosophila melanog
gb.in:AF069986	+ 702.00	964.86	2.2e-45	1385	AF069986 Caenorhabditis eleg
gb.hlg:AF372904	+ 648.50	892.13	2.5e-41	1172	AF372904 Arabidopsis thaliana
gb.hlg:AC010013	+ 632.50	840.14	2.0e-38	34013	AC010013 Drosophila melanog
gb.ba:AE000934	+ 537.50	715.54	1.7e-31	15124	AE000934 Methanobacterium the
gb.pl:SCYJL126W	+ 519.00	711.54	2.9e-31	1307	249401 S.cerevisiae chromosome
gb.pl:AF284571	+ 519.00	709.80	3.6e-31	1591	AF284571 Saccharomyces cerevis
gb.pl:SPC651	+ 502.00	661.73	1.7e-28	25350	AC035570 S.pombe chromosome I
gb.ba:AE004861	+ 500.50	667.48	8.1e-29	10466	AE004861 Pseudomonas aerugin
gb.pr:AF670334	+ 493.00	677.78	2.2e-29	1019	AF260334 Homo sapiens CNA002 m
gb.pl:SCGCD14	+ 493.00	668.77	6.9e-29	2792	Z54149 S.cerevisiae GCD14 gene
gb.ba:STCSLRD	+ 490.00	632.03	7.7e-27	110908	D64002 Synchocystis sp. PCG
gb.pr:BC020620	+ 487.50	670.54	5.5e-29	965	BC020620 Homo sapiens, Nit prot
gb.pr:AF284574	+ 485.50	677.79	7.8e-29	962	AF284574 Homo sapiens, Nit prot
gb.ro:AF284573	+ 481.00	658.94	2.4e-28	1292	AF284573 Mus musculus Nit prot
gb.ba:AB017194	+ 479.00	645.32	1.4e-27	4782	AB017194 Plectonema boryanum c
gb.ro:BC020153	+ 479.00	657.89	2.8e-28	1063	BC020153 Mus musculus, RIKEN c
gb.ba:AP003367	+ 477.50	604.69	2.6e-25	34350	AP003367 Nostoc sp. PCC 7120
gb.ba:AF0005512	+ 476.50	644.73	1.5e-27	3180	AF0005512 Nostoc commune trna-I
gb.sts:G72919	+ 471.00	648.81	8.9e-28	847	G72919 MARC 2849-2850:991933517
gb.ba:AL646071	+ 468.00	596.16	7.6e-25	203050	AL646071 Ralstonia solanaceae

seq\_name: gb.pr:AF069987  
seq\_documentation\_block: 1385 bp mRNA linear PRI 23-JUL-1998  
LOCUS AF069987  
DEFINITION Homo sapiens nitrilas 1 (NIT1) mRNA, complete cds.  
ACCESSION AF069987  
VERSION AF069987.1 GI:3228665  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 1385)  
REFERENCE  
AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Wermuth,P., Rotman,J.H., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Nitrilas and flit homologs are encoded as fusion proteins in  
drosophila melanogaster and caenorhabditis elegans  
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
JOURNAL  
MEDLINE 98337986  
REFERENCE  
AUTHORS 2 (bases 1 to 1385)  
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Wermuth,P., Rotman,J., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Direct Submission  
Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA  
LOCATION/Qualifiers  
1..1385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q21-q22"  
1..1385  
/gene="NIT1"  
77..1060  
/note="NIT1"  
/note="NIT1"  
/codon\_start=1  
/product="nitrilas 1"  
/protein\_id="AAC39907.1"  
/db\_xref="GI:3228666"  
/translation="MGFTTRPHRLSLCPGLRIPLDLSVLCAPRRPAMASSSC  
ELPVAVCOVSTPRKQONFKKCALREARAGLAFAPDPDIPARDPETHLS  
EPFGKLEETVQLARECGMLSLGFFHREGDQWDTOKTYNCHVILNSKAVATIR  
KTHLDCVEIPGGPCWESNTPGPLESPVSTPAKIDLAICYDLRFPESLSIALAQ  
GAEILTPAFSGITGPAHWEVLRARALETCYVVAACGRHHEKRASYSHSVVD  
MGTVVARCEGPGICLARIDINLYLRQLRRHLVFEQHRRPDIYGNLGHPLS"

FEATURES  
source  
gene  
cds  
BASE COUNT 322 a 382 c 357 g 324 t  
ORIGIN  
alignment\_scores:  
Quality: 1748.00 Length: 327  
Ratio: 5.346 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-357-675C-21 x AF069987  
Align seg 1/1 to: AF069987 from: 1 to: 1385  
1 Metleuglypheiethnaryprohisharphelenseuleucy 17  
|||||  
77 ATGCTGGGCTTCATCACAGGCTTCACAGATTCCTGCTTCGTC 126  
17 spfoglneunrgrliepoglnseuervalleucyslaclnproatp 34



```

149 CCAGAACATGTC...TCATCACTTCTCTGGAGCTGCCCTGTGGCT 195
51 ValCysGlnValThrSerThrProAspLysGlnGlnAspPheLysThr 67
196 GTGTGCGAGTATACATCAACACCAACAGCAAGAACTTAAACATG 245
67 salagluLeuValArgGlnValAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
246 TGTGTAGTGGTTCAGAGAGCTGCCAGACTGGTGGCTGGCTGGCTTC 235
84 eubProGlnAlaPheAspPheLeuAlaArgAspProAlaGluThrLeuHis 100
296 TGGCTGAGGATTTGACTTATTTGACAGAAACCTGGCCGAGACATTAATC 345
101 LeuSerGluProLeuGlyLysLeuLeuGluGluLysThrThrLeuAla 117
346 CTGTCCGAACCACTGAATGGGATCTTTGGCCAAATATAGCCAGCTGCG 395
117 aArgGluCysGlyLeuThrProLeuSerLeuGlyLysPheHisGluArgGly 134
396 CAGGGAATGTGGAAATCTGGCTGTCTTGGCGGTTTCCAGAGACCTGGCC 445
134 LAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
446 AAGACTGGGAGCAGAAATCAGAAATCTACAAATGTGATGCTTTGAAAC 495
151 SerLysGluAlaValAlaThrTyrArgLysThrHisLeuGlyAspVal 167
496 AGCAAGAGATCACTAGTGGCCAGTTCAGAGAAACACATCTGTGCGATGT 545
167 LglnIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
546 AGAGATCCAGCTCAGGCGCCGATGAGAAACCACTATACCAAGCTG 595
184 LProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
596 GAGGCACTTTGAGCACCTGTCAAGACCGGCTGGCAGGTGGCTGA 645
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 217
646 GCATATGTTATGACATGCGGTTCCCGAATCTTTGAAATGGCTCA 695
217 naIaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
696 AGCTGGGCGCAAAATCTTACTTATCTTACGCTTTGGATGCTTACAG 745
234 LProAlaHisTrpGluValLeuLeuArgAlaAlaArgAlaIleGluThrGln 250
746 GTCCAGCCCACTGGAGGTGCTGCTGGGCGCGCCCATTTGAATGTCA 795
251 CysTrpValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
796 TGTATGATATAGACAGCAGCGAGTGGACGCCACCATTAACACAGAGC 845
267 aSerTrpGlyHisSerMetValValAspProTrpGlyThrValValAla 284
846 AATTTATGGCATAGCATGTGTGACCCGTGGCGACAGTGTGTGCC 895
284 rGcysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
896 GGTGTCGCGAGGACAGGCTGCTGCTGCGAATTAATGATTCACACTTT 945
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
946 CTACACAGATGGCGCAACACCTGCTGTGTCAGACGACGACAGACTGA 995
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
996 CCTTATGGCAGTCTGGTCACTCACTCTCT 1026
seq_name: gp_ro:BC021634
seq_documentation_block:

```

```

LOCUS      BC021634                1365 bp      mRNA      linear      ROD 22-JAN-2002
DEFINITION Mus musculus, nitrlase 1, clone MGC:13825 IMAGE:4008543, mRNA,
ACCESSION  BC021634
VERSION    BC021634.1 GI:18204912
KEYWORDS   MGC.
SOURCE      house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   Strausberg, R.
            Direct Submission
            Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: CGAP help desk
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbio.org
            Contact: amadan@systemsbio.org
            Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

FEATURES             source
     source
     1..1365
         /organism="Mus musculus"
         /db_xref="locusID:27045"
         /db_xref="taxon:10090"
         /clone="MGC:13825 IMAGE:4008543"
         /tissue_type="Mammary tumor metastasized to lung.
            MMV-ITR/Mnt1 model. Expression driven by an MMV-ITR
            enhancer."
         /clone_lib="NCI_CGAP_Lu30"
         /lab_host="DH10B"
         /note="vector: pCMV-SPORT6"
         59..1030
         /codon_start=1
         /product="nitrlase 1"
         /protein_id="AAH21634.1"
         /db_xref="GI:18204913"
         /translation="MLGFTTRPHOLLCTGYRLTRPVLCQTPRPRTSSSTSWELPL
            VAVCVSTSTPNQENFKTCAELVQDAARIGACIAPAFPIIARNPETLLSPLN
            GDILGOYSQIAREGCIIMLSIGFHEHGDQBNQKLYNCNHYLNSGVSYSRYTHL
            CDVEIPGGPMESNMYTRPGFTLEPPTVYKGLVATCYDMRPPELSIKLAQACAEI
            LTVSARSGSVTPRAHWEYLRRATESSCYVIAAOCGRHHTRA SYGHSWYVDWGT
            VVARCSSEGGICLARIIDLHFLQDMQNLPRVORHRRPDLXSLGHPLS"

BASE COUNT      368 a      334 c      331 g      332 t
ORIGIN
alignment_scores:
    Quality: 1476.00      Length: 327
    Ratio: 4.824          Gaps: 2
    Percent Similarity: 93.578      Percent Identity: 84.098

alignment_block:
US-09-357-675C-21 x BC021634 ..
Align seq 1/1 to: BC021634 from: 1 to: 1365
1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeu 17
|||||

```

```

59 ATGTGGGCTTCATCACCAGGCTCCTCACAA.....CTCCTGTG 99
17 SPProGluLeuArgIleProGluLeuSerValLeuCysAlaGlnProArgP 34
100 TACCGGATACCGATTACTTGAACCCCACTACTTTTACTCAGCCAGGC 149
34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
150 CCAGAACCATGTGC...TCATCAACTTCTGGAGAGCTGCCCTGGGGCT 196
51 ValGlyGlnValThrSerThrProAspGlyGlnGlnAspPheLysThrCy 67
197 GTGTGCCAGTAAATCAATCAACCAACCAAGCAAGCAACTTTAAACATG 246
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlnGlnAlaGlnAlaPheL 84
247 TCGTGAAGTTGGTTCAAGAGGCTGCCAGACTGGGTGCTTGCCTGGCTTTC 296
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
297 TCCCTGAGGCAATTTGACTTTATTTGCAGAAACCCCTGCCAGACATTATC 346
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
347 CTGTCCGAACCACTGAATGGGATCTTTTGGCCCAATATATAGCCAGCTTC 396
117 aArgGluCysGlyLeuThrPheSerLeuGlyGlyPheHisGluArgGlyG 134
397 CAGGGAAATGTGAATCTGGCTCTCTTGGGGCGGTTTCCAGAGCGTGGCC 446
134 InAspTPeGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
447 AAGACGAGGAGCAAGATCAGAAATCTACAAATGTGATGCTTGTGAAAC 496
151 SerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAspAla 167
497 ACCAAGGAGTACAGTAGTGGCAGTTACAGAAACACACTGTGCGATGT 546
167 LeuIlePheProGlyGlnGlyProMetCysGluSerAsnSerThrMetPro 184
547 AAGATCCAGGTCAGGGGCCGATGAGAAACCACTATACCAACCCG 596
184 LyrProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
597 GAGGACACTGTGAGCCACCTGTCAACAGACACCGGCTGCAAGGTTGGTCTA 646
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
647 GCAATCTGTATGACATGCGGTTCCCTGAACTTTCTTGAATGGCTCA 696
217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
697 AGCTGGGGAGAAATACTTACTTATCTTCACGCTTGGATCTGTACAG 746
234 LyrProAlaHisThrPgluValLeuLeuArgAlaArgAlaIleGluThrGln 250
747 GTCACAGCCCACTGGAGGTGCTGCGGGGCCGCCCATGAATCTCAG 796
251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
797 TCCCTATGTAATAGCAGCAGCGAGTGTGAGCCCAACCAAGCAAGACAGC 846
267 aSerTyrGlyHisSerMetValValAspProThrPgluThrValAlaAla 284
847 AAGTTATGGCCATAGCATGTGTGACCCGTTGGGGCAGAGTGTGGCC 896
284 rGCGSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
897 GCTGCTCCGAGGAGCAGGCTCTGCTGCTGTGATGATGATCCCACTTT 946
301 LeuArgGlnLeuArgArgHisLeuProValLpheGlnHisArgArgProAs 317
947 CTACACAGATGCGCCAAACACCTGCTGTGTTCACACCGCAGACCTGA 996

```

```

317 pleuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
997 CCTCTATGGCACTGTGGTCATCCACTCTCT 1027

seq_name: gb_pr:AF069984
seq_documentation_block:
LOCUS AF069984 4079 bp DNA linear PRI 23-JUL-1998
DEFINITION Homo sapiens nitrlase homolog 1 (NIT1) gene, alternatively spliced
product, complete cds.
ACCESSION AF069984
VERSION AF069984.1 GI:3242977
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 4079)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrlase and fhit homologs are encoded as fusion proteins in
drosophila melanogaster and caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
98337986
MEDLINE
REFERENCE
2 (bases 1 to 4079)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES
source
location/Qualifiers
1..4079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q21-q22"
join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3569)
/gene="NIT1"
/note="alternatively spliced"
/product="nitrlase homolog 1"
635..3569
/gene="NIT1"
/gene="NIT1"
635..680
/gene="NIT1"
/note="alternatively spliced"
/number=1
join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3267)
/gene="NIT1"
/note="NIT1; alternatively spliced"
/codon_start=1
/product="nitrlase homolog 1"
/protein_id="FAC39901.1"
/db_xref="GI:3242978"
/translation="MISGFTRRPHRFTSLCPGRITPQLSVLCAOPPRAMATSSSC
ELPLVAQVYTPDPKQNFRTCAELVREARAGACAPLPEAPDFLAPRAPETLHS
EPLGKLLLEETVQLARECGMLSLSGFHEGQDQFOKLYNCHVLNLSGAAVAATYR
KTHLIDVEIPGQGPMESENSTMPGSPLESFVSPAGTIGLAVCYDMKFPRLSLALQA
GAELITPSPAFSGSTGPAHNEVLLRABAIETOCYVVAACGSRHNEKRSYGHSMVVD
PMGTVVARCSESGPGLCLARDLNTLRQLRHHLPFQHRRPDLYGNLHPLS"
681..711
/gene="NIT1"
/note="1A; alternatively spliced"
712..741
/gene="NIT1"
/note="1B; alternatively spliced"
940..1053
/gene="NIT1"
/note="1C; alternatively spliced"
exon
exon
exon

```



```

exon      1288..1383
           /gene="NTT1"
           /number=2
exon      1636..1890
           /gene="NTT1"
           /number=3
exon      2011..2114
           /gene="NTT1"
           /number=4
exon      2321..2454
           /gene="NTT1"
           /number=5
exon      2609..2734
           /gene="NTT1"
           /number=6
exon      3001..3569
           /gene="NTT1"
           /number=7
BASE COUNT 940 a 1082 c 1043 g 1012 t 2 others
ORIGIN

```

```

alignment_scores:
    Quality: 1468.50      Length: 660
    Ratio: 4.505          Gaps: 8
Percent Similarity: 49.394 Percent Identity: 48.939

```

alignment\_block:

US-09-357-675c-21 x AF069984 ..

Align seg 1/1 to: AF069984 from: 1 to: 4079

```

2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
|||||
1289 CTGGGCTTCATCAACAGCCCTCCACAGATTCTCTCTCTGTGTC 1338
18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
|||||
1339 TGGACTCCGGATTACCTCAACCTCAGTACTTGTGCTCAGCCAGGTAC 1388
32 .....
1389 ACGTTTGTGTGTCCTCAGTGCCTGCGACCTAGATGCTCAGTTGTTAA 1438
32 .....
1439 ATGAGATAGTGAGACACAGAGTGTCACTATCCACACATTGATTGCT 1488
32 .....
1489 GAGCCCTACTAGCCCTGGTCAACGTCCCTGTAGAGCATGATCAAAAG 1538
32 .....
1539 GAAGTCCAGCTTCCCTGCTCTCCACTTGACACCCCTTAGCATTAATTG 1588
33 .....ArgPr 34
1589 CTTCCTGTGCTATGAATCTGAGAACTGCGCATGCTGTTCACAGGCC 1638
34 oATGAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuAlaVal 51
|||||
1639 CAGAGCCATGGCTATCTCTCTCTCTCTCTGCGAATGCCCTGGTGGCTG 1688
51 aLcysGlnValIThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
|||||
1689 TGTGCCAGGTAACTCGACGCGACAGACAAGCAAGCACTTTAAACAATGT 1738
68 AlagIleuValIArgIleAlaIaIaArgLeuGlyAlaCysLeuAlaPheLe 84
|||||
1739 GCTGAGCTGTGTCGAGAGGCTGCAGACTGGGTGCCCTGGCTTCTCT 1788
84 uProGluAlaPheAspPheIleAlaIaArgAspProAlaGlnThrLeuHsL 101

```

```

|||||
1789 GCCTGAGGCAATTGACTTCATTTGACAGGGGACCCGACAGAGCGCTACACC 1838
101 euSerGluProLeuGlyGlyLysLeuLeuGlnGluIuYrThrGlnLeuAla 117
|||||
1839 TGTCTGAACCACTGGGTGGGAACTTTGGAAGATACACCCAGCTTACC 1888
117 .....
1889 AGTATCAGGGAATAGCGAGGAGAGGTAGAAATCTTGTTCGACAGTGT 1938
117 .....
1939 CCCTGGGTTGCCAGATATGAGGAGTAGACCTTGAGAGTCAGTGAAGCT 1988
118 .....ArgLcysGlyLeuThrPleuSerLeuGly 127
|||||
1989 TGTCAGTGTCCCTTCCCCCGAGGAATGTGGACTCTGTGCTCTCTGGGT 2038
128 GlyPheHisGluArgGlyGlnAspIleGlnIleThrGlnLysIleTyrAs 144
|||||
2039 GGTTCATGAGCGGTGGCCAAAGACTGGGAGCAGACTCAGAAATCTTACAA 2088
144 nCysHisValLeuLeuAsnSer..Lys.....
|||||
2089 TTGTACAGTGTGCTGCTGAACAGCAAAAGTGAACATTATACCTTTAGCC 2138
152 .....
2139 TGCCCTTCCCATGCTCTCTTACCTAGATCTCCAGAAATGTTTCTCAAC 2188
152 .....
2189 TCTTATTCTTGAACCAAGGATTAGGGGTGTCCTACTTCAGTTCCTA 2238
152 .....
2239 GCCCTATAACTATCTCCTCTGGGAGAGATAGACAGGCTTCTAGAACCA 2288
153 .....GlyAlaValAlaIaThrT 159
|||||
2289 CCAGCACTGATATTCTCTTCTTACTGTAGGGGACATAGGGCCACTT 2338
159 YrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
|||||
2339 ACAGAAACACATCTGTGTACGTAGAGATCCAGGCGGCGCTATG 2388
176 CysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerProValSe 192
|||||
2389 TGTGAAGCAACTCTACCATGCTGGGCCAGTCTTGAGTCACTGTGAC 2438
192 rThrProAlaGlyLysIleGly..Leu.....
|||||
2439 CACACACAGCAGCAAGGTAGAGAGTTGTGAAGATGAGAGGAGGACAG 2488
200 .....
2489 GAATACTTGAATCGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGTA 2538
200 .....
2539 ATGGAATATGACTAGATGCTGTGACAAACAGACAGAGAAACACTACTAG 2588
201 .....AlaValCysTyrAspMetArg 207
|||||
2589 TAGCTGTTTTCATTCAGATTGATGCTGCTGCTGCTATGACATGGG 2638
208 PheProGluLeuSerLeuAlaLeuAlaGlnIaIaGlyAlaGluIleLeuTh 224
|||||
2639 TTCCCTGAACCTCTCTGCGCAATGGCTCAACGTCGAGCAGAGATACTAC 2688
224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisThrPleuVal. 240
|||||

```

```
2689 CTATCCTTCAGCTTTGGATCATTCACAGGCCAGCCCACTGGAGGTAA 2738
241 .Leu..... 241
2739 GATGATGCCCTTTTAAACATAAGGCCCTTTCTTAACCTCACTCCCC 2788
241 ..... 241
2789 CCTTGGCCCTACCAAGTTAAATTCCTCCCTTCCACCTAATGGAAAC 2838
241 ..... 241
2839 TCATTCGCCAGATATTCTCTCATGATAGTTAAATAGTAAATCATCC 2888
241 ..... 241
2889 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGCTGTATATGCC 2938
241 ..... 241
2939 TCACCTGTCACTTCCCACTATTGCTACATGTACTTAAGTGAACACACA 2988
242 .....LeuArgAlaArgAlaIleGluThrGlnCysTyr 252
2989 TCTCATGCCAGGTGTGTGCGGGCCGCTGTATCGAACCACCGTCTA 3038
252 TTAValAlaAlaAlaGlnCysGlyArgHisIleGluLysArgAlaSer 269
3039 TGTAGTGGCAGCAGCACAGTGTGAGCCCAACCATGAGAAAGAGCAGATT 3088
269 TGTAGTGGCAGCAGCACAGTGTGAGCCCAACCATGAGAAAGAGCAGATT 3088
3089 ATGGCCACACATGTGTGTAGACCCCTGGGGAACAGTGTGGCCCGCTGC 3138
286 SerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeu 302
3139 TGTAGTGGCAGCAGCTCTGCTGTGGCCGAATAGACTCAACTATCTGCG 3188
302 GGTGLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeu 319
3189 ACAGTTCGCCGACACCTGCTGTGCCAGCAGCAGCGAGGCTGACCTCT 3238
319 TGTAGTGGCAGCAGCTCTGCTGTGCCAGCAGCAGCGAGGCTGACCTCT 3238
3239 ATGGCAATCTGGGTACCCACTGCTCT 3264
seq_name: gb_hhg:AL354714_4
seq_documentation_block:
WPCOMMENT
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_0 1 110000
AL354714_1 100001 210000
AL354714_2 200001 310000
AL354714_3 300001 410000
AL354714_4 400001 510000
AL354714_5 500001 610000
AL354714_6 600001 699820
Continuation (5 of 7) of AL354714 from base 400001 (AL354714 Homo sapiens chromosome 1)
alignment_scores:
Quality: 1468.50 Length: 660
Ratio: 4.505 Gaps: 8
Percent Similarity: 49.394 Percent Identity: 48.939
alignment_block:
US-09-357-675c-21 x AL354714_4 ..
Align seg 1/1 to: AL354714_4 from: 1 to: 110000
2 LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCysPsr 18
```

*2002*  
*(radio)*  
*Nov*  
*is 2000*

```
103858 CTGGGCTTCATCACACAGGCTCTCCACAGATCTCTGCTGTGCTCC 103907
18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
103908 TGGACTCGGATACCTCACTCACTCACTCACTCACTCACTCACTCACT 103957
32 ..... 32
103958 ACCTTTTGTGTGTCTCACTCACTCACTCACTCACTCACTCACTCACT 104007
32 ..... 32
104008 ATGAGATAGTGAGACACAGAGAGTCACTCACTCACTCACTCACTCACT 104057
32 ..... 32
104058 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGACATGATCAAAAG 104107
32 ..... 32
104108 GAATCCAGCTTTCCTGCTCTCCACTGACCCCTTAGCATTAATTTG 104157
33 .....ArgPsr 34
104158 CTTCCTGTGCTATGAATCTGAAATCTGCTGCTGCTGCTGCTGCTGCTGCT 104207
34 oArgAlaIleAlaIleSerSerSerSerSerSerSerSerSerSerSerSer 51
104208 CAGAGCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 104257
51 aLcysGlnValThrSerThrProAspLysGlnIleAsnPheLysThrCys 67
104258 TGTGCCAGGTAATCATGACGCCAGACACAGACAGCAAACTTAACAATGT 104307
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
104308 GCTGAGCTGCTGTGAGAGGCTGACAGCTGCTGCTGCTGCTGCTGCTGCT 104357
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
104358 GCTTGAAGCATTTGACTTCACTTTCACAGGACCTTGCAGAGACCTTACAC 104407
104358 GCTTGAAGCATTTGACTTCACTTTCACAGGACCTTGCAGAGACCTTACAC 104407
101 eSerGluProLeuGlyGlyLysLeuLeuGlnLysTyrThrGlnLeuAla 117
104408 TGTCTGAACCATGGGTGGGAACTTTTGGAGAATACACCAAGCTTGCC 104457
117 ..... 117
104458 AGGTATCAGGGAATAGCGAGGAGAGGTGAATCTTTGTGGACAGTGT 104507
117 ..... 117
104508 CCCTGGGTGCCAGATATAGAGGTAGAGCTTGAGAGAAGTATGAGAAAGT 104557
118 .....ArgLcysGlyLeuTyrPheLeuSerLeuGly 127
104558 TGTCAAGTGCCTTCCCTCCCAAGGAATGTGAGCTGTGGCTGCTGGGT 104607
128 GlyPheHisGluArgGlyGlnAspTyrPheGlnThrGlnLysIleTyrAs 144
104608 GGTTCATGAGGTGAGCCCAAGACAGACAGACAGACAGACAGACAGACAG 104657
144 nCysHisValLeuLeuAsnSer..Lys..... 152
104658 TGTCTACAGTGTCTCTGTAACAGCAAAAGTGAAGCTTATTAACCTTTAG 104707
152 ..... 152
104708 TGCCTCTCCCATGCTCTTCTACTAGATTCTCCAGATTTGTTCTCAAC 104757
152 ..... 152
```

104758	TCTATTTCCTTGACCACAGAGATTAGGGGTGTCCTACTTCAGTTCC	104807
152	.....	152
104808	GCCATATAACTATCCCTCCCTTGGAGAGATGAAGCAAGGCTCTGAGACA	104857
153	.....	153
104858	CCAGCACTGATATTCCTTCTTCTTACTGTAGGGGACAGTAGGGCCACT	104907
159	YRAGLYTHRHisLeucCysAspValGluIleProGlyGlnIleProMet	175
104908	ACAGGAGACACACTCTGTGTACGTAGAGATTCCAGGGCAGGGGCGCTATG	104957
176	CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSer	192
104958	TGTGAAGCACTCATCACTCCCTGGGCCAGTCTTGAGTACACCTGTGAG	105007
192	rThProAlaGlyLysIleGlyLeu.....	200
105008	CACACCGACGAGCAAGTAGAGATTGTGAAGAGATAGAGGAGGGAACAG	105057
200	.....	200
105058	GAATACTTTGAATCGCAGTAGAGATGAAGAACCCCTAGAGAGGGGTA	105107
200	.....	200
105108	ATGGAATATATCATAGATGCTGTGACAAACAGACAGAAAGCTACTAG	105157
201	.....	201
105158	TAGGCTGTTTTTCATTCAGATTTGGTCTAGCTGTCTGATACATGCGG	105207
208	PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh	224
105208	TTCCCTGAACCTCTCTGGCATTTGGCTCAAGCTGAGCAGACATCTTAC	105257
224	rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal.	240
105258	CTATCCCTTCACGCTTTGGATTCATTCACGCGCCACCCACTGGAGGTAA	105307
241	Leu.....	241
105308	GATGATGCTTTTAAACATAGAAGGCTTTTCTTAACCTCATCTTCCC	105357
241	.....	241
105358	CCTTGGCCCTACACAGTAAATTCCTTCCCCTTCCACCTAATGGGAAAAC	105407
241	.....	241
105408	TCATTTCCCAAGATATTCTCTCATGTGAATAGTAAATAGTAATCATTCC	105457
241	.....	241
105458	TAGCAATTACCAAAATAGTCACATAGGTAGATTGGTCTGAATGTCCC	105507
241	.....	241
105508	TCACCTGTCACTTCCCACTAATTTGCTACATGTACTTAAGTAGACACACA	105557
242	.....	242
105558	TCTCATGCCCAAGTGTGTGCTGGGGGCGCGTCTATGAAACCAAGTGTCTA	105607
252	rValValAlaAlaAlaGlnCysGlyIleArgHisHisGluIleSerTrpGly	269
105608	TGTATGTGCAGACAGACAGTGTGTGAGCCACCATGAGAGAGAGCAAGTT	105657
269	YRGLYHisSerMetValValAspTrpTrpGlyThrValValAlaAspCys	285
105658	ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACATGTGTGGCCCGCTCC	105707

```

286 SerGIgLYPrGlyLeuCysLeuAlaArg11LeaSPLeuSntYTLLeuAr 302
|||||
105708 TCTGAGGGGGCCAGCGCTCTGCTTGGCCGAATGACCTCAACTATCTGCG 105757
302 gGlnLeuArGArgH1sIeuProValPheGlnH1sArgArGProAspLeuT 319
|||||
105758 ACAGTTGGCCCGCACACCTGCTGTGTTCCAGCACCGCAGGCTGACCTCT 105807
seq_name: gp_htg:AL354714_5

seq_documentation_block:
WPCOMMENT
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_0 1 110000
AL354714_1 100001 210000
AL354714_2 200001 310000
AL354714_3 300001 410000
AL354714_4 400001 510000
AL354714_5 500001 610000
AL354714_6 600001 699820
Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome
alignment_scores:
Quality: 1468.50 Length: 660
Ratio: 4.505 Gaps: 8
Percent Similarity: 49.394 Percent Identity: 48.939

alignment_block:
US-09-357-675C-21 x AL354714_5 ..
Align seg 1/1 to: AL354714_5 from: 1 to: 110000

2 LeuGIYPhelIethrArGProProH1sArgPheLeuSerLeuLeuCySPR 18
|||||
3858 CTGGGGCTTCATCACACAGCGCTCTCCACAGATTCCTGTCCTTCGTCTGTC 3907
18 oGlyLeuArg11eProGlnIeuSerValLeuCyAlaGlnPro..... 32
|||||
3908 TGGACTCGGATACCTCACTCTGACTCTGACTTGTGCTCAGCCAGGTAC 3957
32 ..... 32
3958 ACGTTTGTGTGTCCTCAGTGCCTGGCACATTAGATGCTCAGTTGTTAA 4007
32 ..... 32
4008 ATGAGTATGGGAGACACAGAGTGTCACTATCCACACATTGATTGGT 4057
32 ..... 32
4058 GAGCCCTACTAGCCCTGGGTCAACGTGCCCCGTGAAGAGCATGATCAAAAG 4107
32 ..... 32
4108 GAAGTCCAGCTTCCCTGCTCTCCACTTGCACCCCTTAGCATTAATTGG 4157
33 .....ArgPr 34
4158 CTTCCTCTGTCTATGAATAATCTGAGATTCCTGGCTATGCTGTTCAACAGGCG 4207
34 oATGAlaMetAla11eSerSerSerSerCyGlnLeuProLeuValAla1 51
|||||
4208 CAGAGCCATGGCTATCTCTCTTCTCTCTGGAACCTGGCCCTGGTGGGTG 4257
51 alycsglnval1ThrSer1ThrProasplysglnGlnAsnphelysThrCys 67
|||||

```

```

4258 TGTGCCAGTACATCGACGCCGACACAGACAACTTTAAACATGT 4307
68 AlagluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
|||||
4308 GTTGGCTGTTCGAGAGGCTGCACAGCTGGTGGCTGCTGGCTTCT 4357
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHsL 101
|||||
4358 GCCTGAGGCATTTGACTTCATTCACGCGACCTGCAGAGACCGCTACACC 4407
101 euSerGluProLeuGlyLysLeuLeuGluIuTrpThrGlnLeuAla 117
|||||
4408 TGTCTGACCACTGGGTGGAACTTTGSAAGATACACCAAGCTTGCC 4457
117 .....
4458 AGGTATCAGGGAATAGCGAGGAGAGTAGAATCTTGTGGACAGTGT 4507
117 .....
4508 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAGTCAGTGAAAGT 4557
118 ..... ArgGluCysGlyLeuTrpLeuSerLeuGly 127
|||||
4558 TGTCAAGTGTCCCTTCCCGGAGGATGTGACTGTGGCTGTCTGGGT 4607
128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144
|||||
4608 GGTTCATCAGCGGTGGCCAAAGACTGGGACAGCTCAGAAATCTTCAA 4657
144 ncYSHsValLeuLeuAsnSer..Lys..... 152
|||||
4658 TTGTCAAGTGTCTGTGAAACAGCAAGGTGAGACTTATAAACCTTTAGCC 4707
152 .....
4708 TGCCTCTCCCATGCTCTTACCTAGATTCTCAGAAATTGTTCTCAAC 4757
152 .....
4758 TCTTATTTCCTGACCCAAAGATTTAGGGGTGGTCTACTTCAGTTCTTA 4807
152 .....
4808 GCCTATTAATATCTCCTCTGGAGAGAGTAAGCAAGCTTCTAGAAACA 4857
153 ..... GlyAlaValAlaIleAsnThr 159
|||||
4858 CCAGCACTGATATTCCTTCTTACTGTAGGGGCGAGTAGTGGCCACTT 4907
159 YrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
|||||
4908 ACAGGAAGACATCTGTGTGACGTAGAGATTCCAGGCGAGGGCCATATG 4957
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
|||||
4958 TGTGAAGCAACTTACCATGCTGGGCCAGCTTGTGAGTCACTGTCTGAG 5007
192 rThrProAlaGlyLysIleGly..Leu..... 200
|||||
5003 CACACCAAGCAGGCAAGTAGAGTTGTGAAGATGAGGAGGGGAACAG 5057
200 .....
5058 GAATACTTTGACTGGCAGTAGAGATAGAAAGCCCTAAGAGAGGGGCTA 5107
200 .....
5108 ATGGAATATGACTAGATGCTGTGACAAACAGACAGGAAGACTACTAAG 5157
201 ..... AlaValCysTyrAspMetArg 207
|||||
5158 TAGCGTGTTTTCATTCACAGATTGGTCTAGCTGTCTGTATGACATGCGG 5207

```

```

208 PheProGluLeuSerLeuAlaLeuAlaGluAlaGlyAlaGluIleLeuTh 224
|||||
5208 TTCCCTGAACCTCTCTCGCATTTGGGCTCAAGCTGGAGACAGATACTTAC 5257
224 rTyrrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240
|||||
5258 CTATCTTCAGCTTTTGTGATTCATTACAGGCCAGGCCACTGGGAGGTAA 5307
241 .Leu..... 241
5308 GATGATGCCCTTTTAAACATTAAGGGCCTTTCTTAACCTCATCTTCCC 5357
241 .....
5358 CTTGGCCCTACAGTTAAATCTCTCCCTTCCACCTAATGGGAAAC 5407
241 .....
5408 TCATTCGCCAGATATTTCTCTCATGATAGTTAAATAGTAATCATTC 5457
241 .....
5458 TAGGCAATTACCAAAATAGTCACATGGTAGATGGTCTGTATATGCC 5507
241 .....
5508 TCACCTGTCACTTCCCATATTGCTACATGTAATTAACTGAACACACA 5557
242 ..... LeuArgAlaArgAlaIleGluThrGlnCys 252
|||||
5558 TCTCATGCCAGGTTGTTGTCGGGCGGCTGCTATGCAAAACCAAGTCTA 5607
252 rValValAlaAlaAlaGlnCysGlyArgHisIleGluValArgAlaSerT 269
|||||
5608 TGTAGTGGCAGCAGCAGATGTGGAGCCACCATGAGAAAGAGCAAGTT 5657
269 YrGlyHisSerMetValAlaAspProTrpGlyThrValValAlaArgCys 285
|||||
5658 ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGTGCCCTGTC 5707
286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTrpLeuAr 302
|||||
5708 TCTGAGGGGCGAGCCCTCTGCTGCGGAATAGACCTCAACATATCTGCG 5757
302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
|||||
5758 ACAAGTGGCCGACACCTGCTGTCTCCAGCACCGAGGCTGACCTCT 5807
319 YrGlyAsnLeuGlyHisProLeuSer 327
|||||
5808 ATGGCAATCTGGGGTCAACCACTGTCT 5833
5808 ATGGCAATCTGGGGTCAACCACTGTCT 5833

seq_name: gb_hvg:AL590651
seq_documentation_block: 167863 bp DNA linear HTG 18-AUG-2001
LOCUS AL590651
DEFINITION Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL590651.4 GI:13990236
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 167863)
AUTHORS Harrison,E.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 7, 2001 this sequence version replaced gi:13751001.
COMMENT

```

```

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA137A12
----- Summary Statistics
Assembly program: XCAP4: version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 164729 bases at least Q40
Consensus quality: 165486 bases at least Q30
Consensus quality: 166169 bases at least Q20
Insert size: 167063; sum-of-contigs
Insert size: 171878; 3.4% error; agarose-fp
Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality
coverage: 8.11x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 9374 9473: contig of 9373 bp in length
* 9474 29891: contig of 20418 bp in length
* 29892 29991: gap of 100 bp
* 29992 36239: contig of 6248 bp in length
* 36240 36339: gap of 100 bp
* 36340 103784: contig of 67445 bp in length
* 103785 103884: gap of 100 bp
* 103885 114878: contig of 10994 bp in length
* 114879 114978: gap of 100 bp
* 114979 126347: contig of 11369 bp in length
* 126348 126447: gap of 100 bp
* 126448 137764: contig of 11317 bp in length
* 137765 137864: gap of 100 bp
* 137865 163160: contig of 25296 bp in length
* 163161 163260: gap of 100 bp
* 163261 167863: contig of 4603 bp in length.
*
Location/Qualifiers
1. 167863
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_1id="RP11-137A12"
/clone_1id="RPC1-11.1"
1. 9373
/note="assembly-fragment:02360
fragment_chain:1
clone_end:17
vector_side:left"
9474..29891
/note="assembly-fragment:01854
fragment_chain:1"
29992..36239
/note="assembly-fragment:00805
fragment_chain:1"
36340..103784
/note="assembly-fragment:01467
fragment_chain:1"
103885..114878
/note="assembly-fragment:00991
fragment_chain:2"
114979..126347
/note="assembly-fragment:01503
fragment_chain:2"
126448..137764
/note="assembly-fragment:01297

```

```

misc_feature      fragment_chain:2"
137865..163160
/note="assembly-fragment:00117
fragment_chain:2"
misc_feature      163261..167863
/note="assembly-fragment:01267
fragment_chain:2
clone_end:sp6
vector_side:right"
BASE COUNT      43927 a 39935 c 40002 g 43197 t 802 others
ORIGIN
alignment_scores:
Quality: 1468.50      Length: 660
Ratio: 4.505          Gaps: 8
Percent Similarity: 49.394      Percent Identity: 48.939
alignment_block:
US-09-357-675C-21 x AL590651/rev ..
Align seg 1/1 to reverse of: AL590651 from: 1 to: 167863
2 LeuGlyPheileIeThrArgProHisArgPheLeuSerLeuLeuGlyPro 18
|||||
151392 CTGGGCTTCATCACCAGGCTCCTCAGACAGATTCCGTCCCTTGTGTCC 151343
18 oGlyLeuArgIleProGluLeuSerValLeuGlyAlaGlnPro..... 32
|||||
151342 TGGACTCGGATACCTCAACTCCTCAGTACTTGTGTCCAGCCAGGTAAC 151293
32 ..... 32
151292 ACGTTTGTGTGTCTCAGTCTGCTGCGACCTAGATGCTCAGTTGTTAA 151243
32 ..... 32
151242 ATGGATAGTGGAGACACAGGAGTGTCACTACATCCACACATTGATGTT 151193
32 ..... 32
151192 GAGCCCTACTAGCCCTGGGTCACGTCGCTGTAGAGCATGATCAAAAG 151143
32 ..... 32
151142 GAAGTCAGCTTCCCTGCTCTCCACTTGCACCCCTTAGATTAAATTGG 151093
33 ..... ArgPr 34
151092 CTTCCTGTGCTATGAATCTGAGAAATCTGCTATGCTGTTCACAGGCC 151043
34 oArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 51
|||||
151042 CAGAGCATGCTATCTCTCTCTCTCTCTCTGCGAACTGCCCTGGGGCTG 150993
51 aLysGlnValThrSerThrProAspLysGlnGlnAspPheLysThrCys 67
|||||
150992 TGTGCCAGTAGTACATCGACGCGACAGCAACAACACTTAAACACAGT 150943
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
|||||
150942 GCTAGAGCTGTTCAGAGAGCTGCCAGACTGGTGTGCTGCTGCTTCT 150893
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
|||||
150892 GCTGAGGCACTTGTGACTTCATTCGACGGAGCCCTCAGACAGGCTAAC 150843
101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla 117
|||||
150842 TGTCTGAACCACTGGGTGGAAACTTTTGAAGATACACCCAGCTTCC 150793
117 ..... 117

```

```

150792 AGGTATCAGGAAATAGCCAGGAGAGTAGAATCTTTGTGACAGTGT 150743
117 ..... 117
150742 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAAGTACGTAGAGT 150693
118 ..... ArgGluCysGlyLeuTrpLeuSerLeuGly 127
150692 TGTCACTGTCCTTCCCCCAGGGAATGTGACTGTGGCTGTCTTGGGT 150643
128 GtYPheHISgluArgGlyGlnAspTrpGluInThrGlnIleuTyrAs 144
150642 GGGTTCCATGAGGCGTGCCCAAGACTGGGAGCAGACTCAGAAATCTACAA 150593
144 nCysHISValLeuLeuAsnSer..Lys..... 152
150592 TTGTCACTGTCCTGTGACACAGCAAGGTGAGACTTATACCTTTAGCC 150543
152 ..... 152
150542 TGCCTCTCCCATGCTCTTCTACCTAGATTCTCAGAAATTGTTCTCAG 150493
152 ..... 152
150492 TCTTATTTCTTGACCCAGAGATTAGGGGTGCTCTACTTCACTTCTTA 150443
152 ..... 152
150442 GCCTATTAACATCTCTCTTGGGAGAGTAGAAGCCTTACAGACA 150393
153 ..... GlyAlaValAlaIleThr 159
150392 CCAGCACTGATATCTCTTCTTCTTACTGTAGGGGAGTAGTGCCACTT 150343
159 yrfArgIysTrpHISLeuCysAspValGluIleProGlyGlnGlyPromet 175
150342 ACAGGAAGACACTGTGTGTGACGTAGAGATTCAGGGGAGGCGCTATG 150293
176 CysGluSerAsnSerThrmEtpGlyProSerLeuGlnSerProValSe 192
150292 TGTGAAGACACTCTACACATGCTGGGCCAGGCTTGAGTCACTGTGCAG 150243
192 rTrpProAlaGlyLysIleGly..Leu..... 200
150242 CACACCAGCAGCGAGGTAGAGTGTGAAGAGATGAGGAGGGAACAG 150193
200 ..... 200
150192 GAATACTTTGACTGCGAGTAGAGATAGAAAGCCCTAGAGAGGGGCTA 150143
200 ..... 200
150142 ATGGAATATGACTAGATGCTGTGACAAACAGACGAGAGACTACTAAG 150093
201 ..... AlaValCysTyrAspMetArg 207
150092 TAGGCTGTTTTTCATTCAGATTGCTCTAGCTGTCTGCTATGACATGCGG 150043
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
150042 TTCCCTGAACTCTCTGTGGCATTGGCTCAAGCTGGAGCAGAGATACTTAC 149993
224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHISTrpGluVal. 240
149992 CTATCTTCAGCTTTTGGATCCATTACAGGCCACGCCACTGGGAGGTAA 149943
241 ..Leu..... 241
149942 GATGATGCCCTTTTAAACATAAGGCTTTTCTAATCATCTTCCCC 149893
241 ..... 241
149892 CCTTGGCCCTACACGTTAAATTCCTTCCCTTTCACCTAATAGGAAAC 149843

```

```

241 ..... 241
149842 TCATTCOCAGATATTTCTCTCATGATAGTAAATAGTAATCATTC 149793
241 ..... 241
149792 TAGCAATTACCAAAATAGTACAAATGGGTAGATTGGTCTTAATGTCC 149743
241 ..... 241
149742 TCACCTGTCACTTCCACCATTTTGTCTACTGTACTTAACTGAACACA 149693
242 ..... LeuArgAlaArgAlaIleGluThrGlnCysTyr 252
149692 TCTCATGCCACAGGTGTTGTCGGGGCCCGTATGCAACCCAGTGTCTA 149643
252 rValValAlaAlaAlaGlnCysGlyArgHISGluLysArgAlaSer 269
149642 TGTAGTGGCAGCAGCAGTGTGGACGCCACCATGAGAAAGAGCAAGTT 149593
269 yrfGlyHISserMetValValAspProTrpGlyThrValValAlaArgCys 285
149592 ATGGCCACACACATGTGTGTAGACCCCTGGGGAACAGTGTGGCCCGCTGC 149543
286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302
149542 TCTGAGGGGCGCAGGCTCTGCTTGGCCGAAATAGACTCAATCATGTGCG 149493
302 gGlnLeuArgArgHISLeuProValPheGlnHISArgArgProAspLeuTr 319
149492 ACAGTTGCCGCCGACACCTGCTGTTCACGACCGCAGCGACCTGACCTCT 149443
319 yrfGlyAsnLeuGlyHISProLeuSer 327
149442 ATGGCAATCTGGGTACCCACTGTCT 149417
seq_name: gp_pr:AL591806
seq_documentation_block:
LOCUS AL591806 200822 bp DNA linear PRI 30-JAN-2002
DEFINITION Human DNA sequence from clone Rpl1-544M22 on chromosome 1, complete
sequence.
ACCESSION AL591806
VERSION AL591806.16 GI:18476709
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 200822)
AUTHORS Harrison,E.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humney@esanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:17902927.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M3 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP

```

database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP11-544M2.2 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-544M2.2 The true left end of clone RP11-137A12 is at 156538 in this sequence. The true right end of clone RP11-381D2 is at 145015 in this sequence.

Location/Qualifiers

```

US-93C
/organism="Homo sapiens"
/d_xref="taxon:9606"
/chromosome="1"
/clone="RP11-54M22"
/clone_lib="RPCI-11.2"
10395..10397
/misc_feature
(note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
44098..44107
/misc_feature
(note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
59585..59604
/misc_feature
(note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
90797..90883
/misc_feature
(note="Single clone region. Reads generated from a
transposon library derived from a single pUC clone.
Restriction digest data confirm the assembly."
90861..90883
/misc_feature
(note="Sequence from uni-directional dgrp big dye
terminator reads only."
49052 c 46183 g 50372 t
BASE COUNT 55215 a
ORIGIN
alignment_scores:
Quality: 1466.50 Length: 660
Ratio: 4.505 Gaps: 8
Percent Similarity: 49.394 Percent Identity: 48.939
alignment_block:
US-09-357-675C-21 x AL591806 ..
Align seg 1/1 to: AL591806 from: 1 to: 200822
2 LeuGIYPheIIethrArProPHISATyGPhleuSerleucenySPR 18
|||||
172789 CTGGGGCTTATACCCAGGCCCTCTCACAGATTCTGTGTCCTTGTC 172838
|||||
18 oGLYLeuArgIleProGLInleuSerValIeuCYsaIagInPro..... 32
|||||
172839 TGGACTCGGAAACCTCAACTCTCACTACTGTGTGCTACCACCAAGTAAC 172888
|||||
32 ..... 32
172889 ACGTTTGTGTGTCTCCTCAAGCCCTGGCACTTAGAGTCACTTGTGTAA 172938
32 ..... 32
172939 ATGGATAGTGGAGACACAGAGACTGTCAACTATCCACACATTGATTTGCT 172988
32 ..... 32
172989 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAAGCATGATCAAAG 173038
32 ..... 32
173039 GAAGTCACAGCTTTCGTGCCTCTCCACTTGCACCCCTTAGCATTAATTTG 173088

```

33	.....	ArgPr	34
173089	CTTCCSTGTGCTATGAATAATTGAGAAATTCCTGCACCTTGCTGTATGGTTCAACAGGCC		173138
34	OARGLAMEIALAIIeSerSerSerCysGluLeuProLeuValAlaV		51
173139	CAGACCSCATGGGTATCTCTCTCTCTCTCTCTCGGACAAGCCCCCTGGTGGCTG		173188
51	aLCysGlnValThrSerThrProAspLysGlnGlnAsnRheUsthrCys		67
173189	TGTGCCAGGTAACTAACGACGCGCCAGACAAACACAGAACTTTAAACATGT		173238
68	ALAgluLeuValArgGluIuaIlaIarTrgleuGlyAlaCysLeuAlaPhe		84
173239	GCTGAAGTGGTTGCGAGAAGGCTGCCAACTGGGGTGGCTGGCTGGCTTCTCT		173288
84	uProGluIuaIaPheAspRheIleAlaArgAspProAlaGluThrLeuIst		101
173289	GCGTAGGCAATTTGACTTCATTTGCAAGGGAGCCCTGCAAGAGAGCGTACAC		173338
101	eUSerGluProLeuGluGlyLysLeuGlnGluTyrThrGlnLeuAla		117
173339	TGTCGAACCACTGGGTGGGAACCTTTTGAAGAAATACACCCAGCTTGGC		173388
117	.....		117
173389	AGGTATCAAGGAAATAGCAGAGGAGAGATGAATCTTTGTGGACAGTGT		173438
117	.....		117
173439	CCCTGGGTTGCCAGATATGAGGGTAGAGGCTTGAGAAAGTCAGTGAAGAGT		173488
118	.....	ArgGluCysGluLeuThrProLeuSerLeuGly	127
173489	TGTCAGTGTCCCTTCCCCCAGAGGAATGTGGACTGTGGCTGTCTTGGGT		173538
128	GIuPheNIsgLuArGluGlnAspTrgIugInThrGlnLysIleTyraS		144
173539	GGTTTCCATGACGCTGGGCCAAGACTGGGAGCAGACTCAGAAATATTAACA		173588
144	nCySHtSVaIlLeuleuAsnSer..Lys.....		152
173589	TTGTACAGTGTGCTGTAACAGGAAGAGTGAAGCTTTATAACCTTTAGCC		173638
152	.....		152
173639	TGCCCTTCCCACAGCTCTCTTAACSTAGATTCTCCAAATATGTTTCSAAC		173688
152	.....		152
173689	TCTTATTTCTTTCAGACCAAGAGATTAGGGGTGGTCTCTACTTCAGTTCTTA		173738
152	.....		152
173739	GCGTTTAAACTATCTCTCTCTTGGGAGGAGTAAACAAGGCTTTAGACAA		173788
153	.....	GlYAlaValAlaIalThrT	159
173789	CCAGCACTGATATTCCTCTTTCTTACTAGTGAAGGCAAGTAGTGCCACCTT		173838
159	yTArGluYSthrHisLeuCysAspValGluIleProGluGlnGlyPromet		175
173839	AACAGGAAGCACATCTGTGTGACGTAGAAATTTCCAAGGCAAGGGGCGTATG		173888
176	CysGluSerAsnSerThrMetProGluProSerLeuGluSerProValSe		192
173889	TGTGAAGCAACTCAACATTCGCTGGGCCCAAGTCTTAGTCAACTGTCAAG		173938
192	tThrProAlaGluLysIleGly..Leu.....		200
173939	CACACACAGAGGCAAGGTGAGAGTTGTGAAGAGATGAGGAGGAGGAGCAAG		173988

```

200 ..... 200
173989 GAATACCTTTGAACGAGCAGTAGAGATAGAAAGCCCTAAGAGAGGGGCTA 174038
200 ..... 200
174039 ATGGAAATATGACTAGATGCTGTGACAAACAGACGAGAGACTACTAAG 174088
201 ..... AlavalCystTyraSpmctary 207
174089 TAGGCGTTTTCATTCAGATTGCTGTAGCTGTCTGTATGATGACGG 174138
208 PheProGluLeuSerLeuAlaLeuAlaGluAlaGluLeuLeu 224
174139 TTCCCTGAACCTCTCTGCGCATTTGCTCAAGCTGAGACAGATACTTAC 174188
224 TyrProSerAlaPheGlySerLeuPheGlyProAlaHisThrGluVal 240
174189 CTATCCTTAGCTTTTGATTCATTCACAGGCCCGCCACTGGAGGTAA 174238
241 Leu..... 241
174239 GATGATGCGCTTTTAAACATAAGGCCCTTTCTTAACCTCATCTTCCC 174288
241 ..... 241
174289 CCTTGCCCTACACAGTTAAATCCCTTCCCTTCCACTAATGGAAAC 174338
241 ..... 241
174339 TCATTTCCCAAGATATTCTCTCATGAATAGTAAATAGTAATCAATGCC 174388
241 ..... 241
174389 TAGGCAATTACCAAAATAGTCACAAATGGGTAGATTGCTGTATATGCC 174438
241 ..... 241
174439 TCACCTGTACCTTCCCACTATTGCTACATGTAATTAAGTAACACACA 174488
242 ..... LeuArgAlaArgAlaIleGluThrGlnCys 252
174489 TCTCATGCCAGGTGTTGCTGGGGCCGCTGCTATGAAACCCAGTGCCTA 174538
252 rValValAlaAlaAlaGlnCysGlyArgHisGlnCysArgAlaSer 269
174539 TGTAGTGGAGACAGCAGCAGTGTGAGACGCCACCATGAGAAAGACAAATT 174588
269 rGlyHisSerMetValValAspProTrpGlyThrValAlaAlaArgCys 285
174589 ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGTGGCCCTGC 174638
286 SerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArg 302
174639 TGTAGAGGGCCAGGCTGTGCTGCTGCCAATAGACTCAACTATCTGGG 174688
302 GlnLeuArgHisLeuProValPheGlnHisArgArgProAspLeu 319
174689 ACAGTTGGCCGACACACTGCTGTGTTCAGACGACGCGGCTGACCTCT 174738
315 rGlyAsnLeuGlyHisProLeuSer 327
174735 ATGGCAATCTGGGTCCACCACCTGCTT 174764
seq_name: gb_htg:AC084821
seq_documentation_block: 211771 bp DNA linear HTG 29-JAN-2002
LOCUS AC084821
DEFINITION Mus musculus chromosome 1 clone rp23-395h6, WORKING DRAFT SEQUENCE,
1 ordered pieces.
ACCESSION AC084821
VERSION AC084821.24 GI:18390260
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

```

```

SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
1 (bases 1 to 211771)
Deschamps, S., Gu, W. and Roe, B.A.
JOURNAL
Mus musculus BAC Clone rp23-395h6
REFERENCE
AUTHORS
2 (bases 1 to 211771)
Deschamps, S., Gu, W. and Roe, B.A.
JOURNAL
Direct Submission
Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Jan 29, 2002 this sequence version replaced gi:16076995.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKMOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 211771: contig of 211771 bp in length.
1 Location/Qualifiers
1 .211771
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-395h6"
/clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT 55622 a 48688 c 49207 g 58231 t 23 others
ORIGIN
alignment_scores:
Quality: 1105.50 Length: 676
Ratio: 3.893 Gaps: 8
Percent Similarity: 42.012 Percent Identity: 37.722
alignment_block:
US-09-357-675C-21 x AC084821 ..
Align seg 1/1 to: AC084821 from: 1 to: 211771
29 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysG1 45
|||||: ||||||:|||||: ||||||:|||||: ||
975 TGTGTATTCACAGGCCGACGAAACCATGTCC...TCTATCACTTCCGGGA 1021
45 uLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGln 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1022 GCTGCCCGCTGGTGGTGTGTCAGAGTACATCAACACCAACAAAGCAAG 1071
62 lAsnPhelysThrCysAlaGlnLeuValArgGluAlaAlaArgLeuGly 78
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1072 AGAACTTTAAACATGTGTGATGTGTCAGAGGCTGCCAGACTGGGT 1121
79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspR 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1122 GCTTGCTGGCCCTTTCTGCTGAGGCAATTTGACTTTATTCACGAAACCC 1171
95 oAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuLeuGluG 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1172 TCCGCGACATTAACCTCTCGCAACCACTGAATGGGATCTTTTGGGCC 1221
112 lutyThrGlnLeuAla..... 117
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



[illegible]

seq\_documentation\_block: 4481bp DNA linear ROD 23-Jul-1998

LOCUS AF069985

238 ..... 238

2172 TCCAGCCTTACGTGATGATGACACAGGTTCAAGCCAGCATGCTTGAC 2221

238 ..... 238

2222 TAAAAAGAAAACAAATAGACGTGGGACGTGGGGTGCACACTTAA 2271

238 ..... 238

2272 TCCAGACTCAGGAGGACAGAGCAGGTGATCTCTAGATTGAGCCAG 2321

238 ..... 238

2322 CCTGGCTCAGAGTGGTTCCAGACAGCCAGGGCTACACAGAAAAC 2371

238 ..... 238

2372 CTGTCTCAAAAACAAAAAAGCAACCCACACTTCAAGAGTGGGGCATG 2421

238 ..... 238

2422 TGGTGATGCTCTTTCATCTCAGCACTGTGAGCAAAAGCAGCAGATCTC 2471

238 ..... 238

2472 TGTGGTCTAAGCCAGCCAGGTTACACAGGAGACCTTATCTCAAAAGT 2521

238 ..... 238

2522 ACAAAACACAACACACTCTCTTCTTACTTGTATTTCTCTTTCATCT 2571

238 ..... 238

2572 GGTGAGAAATTCATTCTTAGTAGTGTGATCATGTAAAGATTAG 2621

238 ..... 238

2622 TTATTAACTAGTTGGTAGTGAGAAAGATTGGCCACAGTCCCGCTT 2671

238 ..... 238

2672 CTGTGCTACTGCCATTAAACGCTGTATGTTCTGGAGCAACGCACTCT 2721

239 ..... Glu.ValIleuLeuArgAlaArgAlaIleGluThrGlnCysTyrVa 253

2722 CATCTCAGGTGCTGCTGCGGGCCGCGCCACTGAATCTCAGTGTATGT 2771

253 ValAlaAlaAlaGlnCysGlyArgHisHisGlnLysArgAlaSerTyrG 270

2772 AAATGACGACGCGCAGTGTGAGCCACCAATGAACAACAGCAAGTATAG 2821

270 LysHisSerMetValIleAspProTrrpGlyThrValIleAlaArgCysSer 286

2822 GCATGACATAGTGGTTGACCCGTTGGGGCAAGAAGTGTGGCCGCTC 2871

287 GluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgG 303

2872 GAGGAGACACAGCCCTCTGCTGCTGCAATTATCTCCACTTTCACAACA 2921

303 nIeuArgArgHisIleuProValIleHisHisArgArgProAspLeuTyrG 320

2922 GATGCGCCACACCTGCTGTGTTCCAGACACCGCAGACTGACCTTATG 2971

2972 GCAGTGTGGGTCAATCACTCTCT 2994

seq\_name: gb\_ro:AF069985

DEFINITION	Mus musculus nitrilase homolog 1 (Nlt1) gene, alternatively spliced product, complete cds.
ACCESSION	AF069985
VERSION	AF069985.1
KEYWORDS	GI:3242979
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 4481)
TITLE	Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tilib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
JOURNAL	Direct Submission
REFERENCE	Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
AUTHORS	Location/Qualifiers
TITLE	1..4481
FEATURES	/organism="Mus musculus"
SOURCE	/db_xref="taxon:10090"
	/chromosome="1"
	/map="1q21-q23"
MRNA	join(541..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3944)
	/gene="Nlt1"
gene	/note="alternatively spliced"
exon	/product="nitrilase homolog 1"
	541..3944
	/gene="Nlt1"
	541..606
	/gene="Nlt1"
	/note="alternatively spliced"
CDS	/number=1
	join(605..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3658)
	/gene="Nlt1"
	/note="alternatively spliced"
	/codon_start=1
	/product="nitrilase homolog 1"
	/protein_id="AAC40184.1"
	/db_xref="GI:3242980"
	/translation="MGFTTRPHOLCTGYRLRIPVLCTOPRRPTSSSTSWELPLVAVCVSTPNKQENFKCAELVQEARAGACAPLEAFDFIARNPAETLLSEPLNGLDGVYSQIAREGIGIMLSIGFHEGQWQENOKLYNCHVILNSGSVAARKTHLCDVEIRQGGPMRESNMYTKPGGTLPEPRVYKRPARKVGAICTYDKRFPGLSLKQACAEI LTPSAFSGVTGPAAHEVILNRAALIESQCTVAAACGGHHETRASVGSVVDPWGT VVARSGEGPGLARIDILHFDQMRQHLPEFQHRRPDLXSLGHPLS"
exon	863..986
	/gene="Nlt1"
	/note="alternatively spliced"
exon	/number=2
	1177..1263
	/gene="Nlt1"
exon	/number=3
	1638..1889
	/gene="Nlt1"
exon	/number=4
	2015..2118
	/gene="Nlt1"
exon	/number=5
	2362..2495
	/gene="Nlt1"
exon	/number=6
	2626..2751
exon	/gene="Nlt1"

[illegible]



```

*      *      1      11373: contig of 11373 bp in length
*      11374      11473: gap of unknown length
*      11474      43477: contig of 32004 bp in length
*      43478      43577: gap of unknown length
*      43578      126526: contig of 82949 bp in length
*      126527      126626: gap of unknown length
*      126627      213203: contig of 86577 bp in length.
FEATURES
      source
          1..213203
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /chromosome="1"
              /clone="rp23-191a19"
              /clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT      57858 a 50465 c 49150 g 55425 t      305 others
ORIGIN
alignment_scores:
      Quality: 1083.00      Length: 676
      Ratio: 3.813      Gaps: 9
Percent Similarity: 42.012      Percent Identity: 37.574
alignment_block:
US-09-357-675c-21 x AC087229 ..
Align seg 1/1 to: AC087229 from: 1 to: 213203
29 CysAlaGlnProArpProArpGlnAlaMetAlaIleSerSerSerCysG1 45
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48642 TGTGTATTCACAGGCCACAGACCATGCTCC...TCATCAACTTCCTGGCA 48688
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 uleuProLeuValAlaValCysGlnValThrSerThrProAspLysGln 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48689 GCGGCCCTGGTGGCTGTGTGCCAGGTAACATCAACACCAACAAGCAAG 48738
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 lnaAspMetLysThrCysAlaGlnLeuValArgGlnAlaAlaArgLeuGly 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48739 AGCAACTTTAAACATGCTGTGATGCTTCAAGAGGCTGCCACACATGGGT 48788
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsp 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48789 GCTTGCCTGGCTTCTGTGCTGAGCAATTTGACTTTATTCACAGAAACC 48838
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 oAlaGlnThrLeuHisLeuSerGluProLeuGlyLysLeuLeuGlnG 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48839 TGGCGAGACATTCCTCTGTCCGAAACACATGAGATGGGATCTTTGGGCC 48888
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
112 lntYThrGlnLeuAlaArg..Glu..... 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48889 AATATAGCCAGCTTGGCAGGTACAAAGGTAGGGTTGGGGGGGGGGTT 48938
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 ..... 119
48939 GTGTGTGTAGAACTCCCTGTGAGCAGTGTCTCGATTTGCCACAGAGG 48988
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 .....CysGlyLe 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48989 GGTAGAGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 49038
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 uTPLeuSerLeuGlyGlyPheHisGlnArgGlyLnaSPTrpGlnGln 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49035 CTGGCTGTCTGTGGGGGTTTCCACGAGCGTGGCCAAAGACTGGAGCAGA 49088
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 hrcGlnLysIleTyraSncYHisValLeuLeuAsnSerLysGlyAla..V 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49085 ATCAGAAATATCAATATGTATGTCTTTTGAACAGCAAGGGGTGAGAGC 49138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 a1..... 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49139 TCATAGCCCTTAGTTGCCCTTCCAGAGCTTCTTACCTTTCTCAACT 49188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 ..... 155

```

```

49189 CAGATCTTCAGCATTTCTTTCACACTGTGGCTTGGGGTTTATGCT 49238
155 ..... 155
49239 TCAGTCTCAGCTACAGATTACTTTGGGAGAAAGCTAAGCTGGACT 49288
155 ..... 155
49289 TGTAGATGTCTACTGATATTCTCTCTCTCTCTCTCTCTCTCTCT 49338
156 .....ValAlaThrTyra 160
49339 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 49388
160 gLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysG 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49389 GAAGACACATCTGTGGATGTAGAGATCCAGGTCCAGGGGCCATGAGAG 49438
177 luserAsnSerThrMetProGlyProSerLeuGlnSerProValSerThr 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49439 AAGCAACTATATACCAAGCTGAGAGCAGCTCTGTGAGCACCCTGTCAAGACA 49488
194 ProAlaGly..Lys..... 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49489 CCGCTGGCAAGGTGGAGATTGGAAAGAAATGGAGACTATGTGAAAG 49538
197 ..... 197
49539 TGGTAAACCCCCCATTTGGTGCAGGCTAATGAAGCAGTACAGTTCTGT 49588
198 .....IleGlyLe 200
49589 TGAATGGAGCAGGAGAAATAGTGGGTGTGTCTCATTTACAGTTGGTCT 49638
200 uAlaValCysTyraSPMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49639 AGCAATCTGTATAGACATGGGTCCCTGAACTTTTGAATTTGGTGCCT 49688
217 lnaAlaGlyAlaGluIleLeuThrTyraProSerAlaPheGlySerIleThr 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49689 AAGCTGGGGCAGAAATACTTACTTATCTTCACAGCTTTGGATGTGTACA 49738
234 GlyProAlaHisTrp..... 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49739 GGTCCAGCCCACTGGGAGGTAAAGAAATGCCCTTCATACATATAGGCC 49788
238 ..... 238
49789 TCCTTAACCTTACTTTTATAGTGTGGAGCAGAGTGGTACAAACTTAAT 49838
238 ..... 238
49839 CCAAGCACTTCAGTGAATAGCCACAGGTTCAAAAGCAGATGCCTTGACT 49888
238 ..... 238
49889 AAAAAAAGAAACAAATAGAACTGGGAGTGGGGGTGCACACCTTTAAT 49938
238 ..... 238
49939 CCCAGCACTCAGAGAGCAGAAAGAGTGTATCTGATTTGAGGCCAGC 49988
238 ..... 238
49989 CTGGTTACAGAGTGGGTTCCAGAGACAGCAGGGCTACACAGAAACC 50038
238 ..... 238
50039 TGTCTCAAAAAAAGCAAAAGCAACCCACACCTTCAAGAGTGGGGCATGGT 50088
238 ..... 238

```



```
* 38421 43244: contig of 4824 bp in length
* 43245 43344: gap of unknown length
* 43345 48228: contig of 4884 bp in length
* 48229 48328: gap of unknown length
* 48329 54595: contig of 6267 bp in length
* 54596 54695: gap of unknown length
* 54696 58551: contig of 3856 bp in length
* 58552 58651: gap of unknown length
* 58652 62204: contig of 3553 bp in length
* 62205 62304: gap of unknown length
* 62305 66587: contig of 4283 bp in length
* 66588 66687: gap of unknown length
* 66688 70929: contig of 4241 bp in length
* 70929 71028: gap of unknown length
* 71029 75627: contig of 4599 bp in length
* 75628 75727: gap of unknown length
* 75728 78272: contig of 2545 bp in length
* 78273 78372: gap of unknown length
* 78373 80301: contig of 1929 bp in length
* 80302 80401: gap of unknown length
* 80402 83675: contig of 3274 bp in length
* 83676 83775: gap of unknown length
* 83776 87173: contig of 3398 bp in length
* 87174 87273: gap of unknown length
* 87274 90196: contig of 2923 bp in length
* 90197 90296: gap of unknown length
* 90297 92065: contig of 1769 bp in length
* 92066 92165: gap of unknown length
* 92166 94612: contig of 2447 bp in length
* 94613 94712: gap of unknown length
* 94713 98755: contig of 4043 bp in length
* 98756 98855: gap of unknown length
* 98856 102543: contig of 3688 bp in length
* 102544 102643: gap of unknown length
* 102644 105833: contig of 3190 bp in length
* 105834 105933: gap of unknown length
* 105934 108499: contig of 2566 bp in length
* 108500 108599: gap of unknown length
* 108600 111672: contig of 3073 bp in length
* 111673 111772: gap of unknown length
* 111773 114662: contig of 2890 bp in length
* 114663 114762: gap of unknown length
* 114763 116145: contig of 1383 bp in length
* 116146 116245: gap of unknown length
* 116246 118147: contig of 1902 bp in length
* 118148 118247: gap of unknown length
* 118249 120579: contig of 2332 bp in length
* 120580 120679: gap of unknown length
* 120680 122618: contig of 1939 bp in length
* 122619 122718: gap of unknown length
* 122719 124539: contig of 1821 bp in length
* 124540 124639: gap of unknown length
* 124640 126651: contig of 2012 bp in length
* 126652 126751: gap of unknown length
* 126752 128460: contig of 1709 bp in length
* 128461 128560: gap of unknown length
* 128561 130592: contig of 2032 bp in length
* 130593 130692: gap of unknown length
* 130693 132864: contig of 2172 bp in length
* 132865 132964: gap of unknown length
* 132965 134881: contig of 1917 bp in length
* 134882 134981: gap of unknown length
* 134982 136682: contig of 1701 bp in length
* 136683 136782: gap of unknown length
* 136783 139138: contig of 2356 bp in length
* 139139 139238: gap of unknown length
* 139239 140757: contig of 1519 bp in length
* 140758 140857: gap of unknown length
* 140858 142558: contig of 1701 bp in length
* 142559 142658: gap of unknown length
* 142659 144303: contig of 1645 bp in length
* 144304 144403: gap of unknown length
* 144404 145490: contig of 1087 bp in length
```

```
* 145491 145590: gap of unknown length
* 145591 147171: contig of 1581 bp in length
* 147172 147271: gap of unknown length
* 147272 148639: contig of 1368 bp in length
* 148640 148739: gap of unknown length
* 148740 150295: contig of 1556 bp in length
* 150296 150395: gap of unknown length
* 150396 151505: contig of 1110 bp in length
* 151506 151605: gap of unknown length
* 151606 152813: contig of 1208 bp in length
* 152814 152913: gap of unknown length
* 152914 153423: contig of 2510 bp in length
* 153424 155523: gap of unknown length
* 155524 156880: contig of 1357 bp in length
* 156881 156980: gap of unknown length
* 156981 158045: contig of 1065 bp in length
* 158046 158145: gap of unknown length
* 158146 159255: contig of 1110 bp in length
* 159256 159355: gap of unknown length
* 159356 160417: contig of 1062 bp in length
* 160418 160517: gap of unknown length
* 160518 161644: contig of 1127 bp in length
* 161645 161744: gap of unknown length
* 161745 163180: contig of 1436 bp in length
* 163181 163280: gap of unknown length
* 163281 164483: contig of 1203 bp in length
* 164484 164583: gap of unknown length
* 164584 166039: contig of 1456 bp in length
* 166040 166139: gap of unknown length
* 166140 167290: contig of 1151 bp in length
```

```
alignment_scores:
  Quality: 1059.00      Length: 835
  Ratio: 3.507          Gaps: 11
  Percent Similarity: 36.168  Percent Identity: 31.976
```

alignment\_block:

US-09-357-675c-21 x AC105589 ..

Align seg 1/1 to: AC105589 from: 1 to: 178418

```
2 LengthPheniletharprophorohisargphenylserleucyspr 18
|||||
39095 CTGGGCTTCATCCACGAGCCTCCATCACTCCCTTGCTGTCAG 39144
18 oglyleuargileproglinleuvalleucysalaginpro..... 32
|||
39145 CGGATACCGATTACCTCAATCTCAGTACTTGTACTCAGCCAGGTAC 39194
32 ..... 32
32 ..... 32
39195 ATATTTATGTTCTTGTGTACGACGAGTGTGTAAGTCATGTTATTA 39244
32 ..... 32
39245 CCCGAATAGTAGACCTGCATCCACCAACATTAAATGCGTACTCCAG 39294
32 ..... 32
39295 TAGCTTAGGCTTATTAAATGCTCTGGAAGATGAAAAACAAGTCGTA 39344
32 ..... 32
39345 TTGTGGGACACACCTCTATCCACGACACTCGGGAACAAGCAGAGGAT 39394
32 ..... 32
39395 CAACTAGAGATTAGATCATCCCGCTACATAGAAAGTCTGCCGTGGGC 39444
32 ..... 32
39445 TCACAAAAGACCCGGTCTCAAGGGAGGATAAATCAGAAAAAAG 39494
```

```
32 ..... 32
39495 TAGTTCCGGCTCCACCCCACTTGAGCTCCCAAGTTCTCCGGGCA 39544
33 ..... ArgProArgAl 36
39545 CCGGAGAAATCCTGAGAAACTGCTGTGTCATTTCACAGGCCAGAGC 39594
36 AneTAlaIleSerSerSer...SerCysGluLeuProLeuValAlaValC 52
39595 CATGGCAGACTCTCATCGACTTCTGCGAGCTGCCCTGCTGGCTGTGT 39644
52 YsGlnValThrSerThrProAspLysGlnGlnAsnPhelYsthrCysAla 68
39645 GCCAGGTAACTACACCAACCAACAGCAAGAACTTTAAACGTGTGCT 39694
69 GluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhelLeuP 85
39695 GACCTGGTTCAGAGGCTACAGACTGGGGCTTGCCTGGCTTCTGCC 39744
85 OGluAlaPheAspPheIleAlaIleArgAspProAlaGluThrLeuHisLeuS 102
39745 TGGAGCATTTGACTTTATTCACGAAATCCTGCCAGACATTTACTCTGT 39794
102 erGluProLeuGlyGlyLysLeuLeuGluGluYrThrGlnLeuAla... 117
39795 CCGAACCACTGGATGGGACCTTTTGGGACATATAGCCAGCTTGCAGC 39844
117 ..... 117
39845 TACAAGGTAGGGTGGGGTGGGGGAGGGGTANAATCCCTGTTGGGAG 39894
117 ..... 117
39895 TGTTCCTGATGCCACATAGAGGGGTAGAGTTGTCAGTCCCTCCGCC 39944
118 ..Arg..GluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgI 133
39945 CTCAGGGAAATGTGAATCTGTGCTGCTTGGTGCTTCCACCAACCTGG 39994
133 yGlnAspTrpGluGlnThrGlnLysIleTyraAsnCysHisValLeuLeuA 150
39995 CCCAGATGGGACAGAAATCCGAAATCTACATTTGCTCGGGCTTCTGA 40044
150 sn..SerLys..... 152
40045 ACAACCAAGGTGAGACATTCCAATCCTTCAACTGGCTTCTCTAAGTC 40094
152 ..... 152
40095 ATCTAACTTCTCTCACTTAGAACTTAACTTTTCTCACTCTGGGCT 40144
152 ..... 152
40145 TTAGGGGTTTTTTTGTGTTTGTGAATGCTTCTGTTCTTAGCCTAAAA 40194
152 ..... 152
40195 TTACTCTTGGGAGAGAGCTAAGCTGAGCTCCAGTGTCTGATAGTCAGC 40244
153 ..... GlyAlaValAlaIleThrTyraIleGlysthrHisL 164
40245 ATTTTCTCTTACTTTAGGATCAGTACGTGGCCAGTTACAGGAAAGCACATC 40294
164 euCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
40295 TATGTATGTAGAGATCCAGGTCAGGGGCTATGAGAGAAAGAACTAT 40344
181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLy 197
40345 ACCATGCTGTGATACGCTTGTGAACCACTGTCAAGACACAGAGAGCAA 40394
197 sIleGly..Leu..... 200
40395 GGTAGAGTTTGAAGAAATAGGAATATGTTGAAAAAGTTGAACATC 40444
200 ..... 200
40445 TCCCTTGTGTAGATGATGATGAGTAGTATTCAAATGTTAGTTGACAG 40494
201 ..... AlaVal 202
40495 GACAGGAGATGATGAGGTGTTGTTCTCGTTTCAGGGTGGTCTAGCAATC 40544
203 CysTyraSpmEtaArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaI 219
40545 TGTATGACATGGCGTCCCTGAACTTCTCTGAATGGGCTCAAGCTGG 40594
219 yAlaGluIleLeuThrTyraProSerAlaPheGlySerIleThrGlyPro 236
40595 GCCAGAAATACTTACTATATCTTCCAGCCTTGGATGTGTACCGGCCAG 40644
236 IaHisTrpGluVal..... 240
40645 CCCACTGGGAGTAGAGAAATGCCCTTTCATPAACATATAGCCCTCTTA 40694
240 ..... 240
40695 ATCTTATTTTATGATAGTGGGAGACAGATAGCACAAACCTAACTTAAGC 40744
240 ..... 240
40745 ACTTCAGGGGATAGCCACAGGTTCAAGCCAGCATAGTATGCTCGAGT 40794
240 ..... 240
40795 AAAACAGAAACAAATAGAGCTGGGACAGTGGGGGTGCACACCTTTAAT 40844
240 ..... 240
40845 CCCAGCCCTCAGGAGGAGCAAGCAAGTGTGAGTTCAAGACAGCT 40894
240 ..... 240
40895 GGTCTACAGATAGTTCAGGACAGCCAGGCTACACAGAGAAACGGTC 40944
240 ..... 240
40945 TCAAAACAAACAAACAAACAAACAAAGTAACCCACACCAACCCAAA 40994
240 ..... 240
40995 ATAGTGTGATGTGTGTGTGTCTTTTATCTCCGACAGTAAAGCAGA 41044
240 ..... 240
41045 GGCOCGAGATCTCTGTGGGTCCAGGCCAACAAAGTTATACAAGAGAGAC 41094
240 ..... 240
41095 CTTATATAAAAAATACAAAAACAAACACACCCCTCCCTCTTACTTGA 41144
240 ..... 240
41145 AGGATTTTCTCCCTCCGATCTGGTGAGGAAATTCGTTCTTAGATAGG 41194
240 ..... 240
41195 TGTAGTCATGTATAGTTAGTTACTATCTAGTATGATCAATATGCAAT 41244
240 ..... 240
41245 GAGTAGATGGGCCACAGTGTGCGTTCCTGTCACCTGGCTATTAATG 41294
241 ..... Leu..LeuArgAla 244
41294 ..... 244
```

41295	CNGTATGTTCCCTGAGTAGACGAGCTGTCAACCTCAGTGCGTCGGCGG6GC	41344
245	ArgAlaIIEgluThrGlnCysTyValValAlaAlaAlaGlnCysGlyAr	261
41345	CGGGCATTTGAGTCTCAGTGTCTAAGTAAGACAGCACAGTGTGGAGC	41394
261	GHTSHSGILuYSArgAlaSerTyGLynHisSerMetValAlaAspPro	278
41395	CCACCATTGAGACAGAAGCAAGTACGGCCACAGCATGTGTGGACCCCT	41444
278	rpglYthrValValAlaAlaArgCysSerGluGlyPProGlyLeuCysLeu	294
41445	GGGGCCACAGTGTGGTGGCCACTGTTCGAGAGGACACAGGCCCTTGGCT	41494
295	ArgLIAspLeuAsnTYrLeuArgGlnLeuArgArgHISLeuProValAl	311
41495	CGAATTGATCTCCACTTTCTGCACACAGATGGCCACACTTGCTGTGT	41544
311	eGIInHSArgArgProAspLeuTYrGlyLysLeuGlnHisProLeuSer	327
41545	CCAGCATCCGACAGCTGACCTTTATGTGCAAGTTC...CTGCCACTCTCC	41590

```
seq_name: gb_ov:AF284575
seq_documentation_block:
LOCUS       AF284575                1214 bp    mRNA           linear   VRT 22-JUL-2000
DEFINITION  Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.
ACCESSION   AF284575
VERSION     AF284575.1 GI:9367117
KEYWORDS
SOURCE      African clawed frog.
            Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
            Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 1214)
            Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
            Bieganski,P., Pekarsky,T., Croce,C.M. and Brenner,C.
            Crystal structure of the worm NitFhit Rosetta Stone protein reveals
            a Nit tetramer binding two Fhit dimers
            Curr. Biol. 10 (15), 907-917 (2000)
JOURNAL     20414396
MEDLINE     2 (bases 1 to 1214)
AUTHORS     Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
            Bieganski,P., Pekarsky,T., Croce,C.M. and Brenner,C.
            Direct Submission
            Submitted (05-JUL-2000) Kimmel Cancer Center, Thomas Jefferson
            University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA
FEATURES
SOURCE
gene
CDs
1..1214
/db_xref="taxon:8355"
1..1214
/gene="Nit1"
180..1046
/gene="Nit1"
/note="Nit1, similar to the Nit domains of Caenorhabditis elegans NitFhit and Drosophila melanogaster NitFhit"/codon_start=1/product="Nit protein 1"/protein_id="AA87104.1"/db_xref="GI:9367118"/translation="MAGAKPIIAVCOMSTSDPEKEKFAFCSSRLIRAAGRACWFLPEARDTIGSTLEPTLSAESDHGDTTORITQTLAEQCLWLISLGFEHKGNWDDORTSNCHVVNDNTGHIVSYVRKAHLDPVDLDQNVSLRESSSTLPGLALIRPSPKGIGLVNCHDIPERFSLAQQGAELTPFAFTLTGLAIHEVLLRAFALETQCYYVAQAQTDRIENERTSYGHAMVYDPMGLIVIGCCQEGTGICAYIDIPMYRVRBDMVWHRRTDLYKISFNPNPD"
```

alignment\_scores:

BASE COUNT	281 a	305 c	345 g	282 t	1 others
ORIGIN					

	Quality:	968.50	Length:	307
	Ratio:	3.905	Gaps:	1
	Percent Similarity:	80.782	Percent Identity:	57.980
<p>alignment_block:</p> <p>US-09-357-675C-21 x AF284575 ..</p>				
<p>Align seg 1/1 to: AF284575 from: 1 to: 1214</p>				
16	leucySPProGlyLeuAlaIleProGlnLeuSerValLeuCysAlaGlnPr	32		
144	CTGTCTCCGGATATGTGGGCTATGAGAACTGAGTGCATG.....	182		
32	okgrProAlaGlnMetAlaIleSerSerSerSerCysGlnProLeuV	49		
183	.....GCTGGTGGCCACAGCCCTTGA	204		
49	alAlaValCysGlnValThrSerThrProAspIlySGlnIAsnPhel	65		
205	TTGGCCGTGGCCAGATGACTTCAACCTCTGTAAAGAAATTTGCC	254		
66	ThrCysAlaGlnLeuValArgGlnAlaAlaArgLeuGlyAlaCysLeu	82		
255	ACGTGTTCGGCGCTGATCCGGGAGGCTCGGGCGCTGCCTTCATGCT	304		
82	arHeuProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThr	99		
305	GTTCCTGCCGGAAGCTTTGACTATATCGGGGCGAGCATTTGAGACAGCC	354		
99	eunHisSerGlnProLeuGlyIlyLeuLeuGlnIlyThrGln	115		
355	TDAGTCTGAGTCTGATCTCATACATGGGACACATTTCAGCGTTACACCAA	404		
116	LeuAlaIArgGlyCysGlyLeuThrLeuSerLeuGlyIlyPheHisGlu	132		
405	CTCGCCAGGAGTGTGGGCTCTGGCTTTCCCTGGGGGATTTTCATGGAA	454		
132	gGlyIAsnSprGlnIugIlnThrGlnIlyIleTyrAsnCysHisValLeu	149		
455	AGGACCACACTGGGACAGGACGACCAACCATTTCCAAATTCACAGTGGTTG	504		
149	eunAsnSerIlySGlyAlaValAlaThrTyrArgIlyThrHisLeu	165		
505	TGGACAAACAGAGGCACATAGTATCGGTGATCCGCAAGCTCACCTGTT	554		
166	AspValIleGlnIleProGlnGlnGlyProMetCysIlySerAsnSerThr	182		
555	GACGTAGACTTGCAGAAATGGAGTCTACATGAGAGAGCAAGTTCCACCT	604		
182	ProGlyProSerLeuGlnSerProValSerThrProAlaGlyIlyIleG	199		
605	CCCCGGACACAGACTTATTCGCCCCCATTCACATCTCCAGCAGGAAAGTTG	654		
199	IlyLeuAlaValCysTyrAspMetArgPheProGlnLeuSerLeuAlaLeu	215		
655	GCCTGGGGGTGTGTTACACACCTCCGCTTCCAGAAATTCCTCTGGGCTTG	704		
216	AlaGlnAlaGlyAlaGlnIleLeuThrTyrProSerAlaPheGlySer	232		
705	GGCCAAACAGAGGAGAACTTCTCATCTTACCTCTTCCCTTCCACCTC	754		
232	ethrGlyProAlaHisThrGlnValIleLeuLeuArgAlaIleGlu	249		
755	TACTGTGTGGCAGCTTGGGAGGTGTGCTGAGAGCCCGGTCATGAGAA	804		
249	hrcIlyCysTyrValValAlaAlaAlaGlnCysGlyArgHisIlyGly	265		
805	CCCAAGCTCTAGTGTGCTGACGCGCACAGACAGCAACAAATGAGAA	854		
266	ArgAlaSerTyrGlyIlyHisSerMetValValAspProIlyThrVal	282		
855	AGGAGCTCTTATGGCAGCGATATGTGTGTAGACCCGTGGGGGCTGGCAT	904		



[illegible]





**THIS PAGE BLANK (USPTO)**



Quality: 1743.00 Length: 326  
Ratio: 5.347 Gaps: 0  
Percent similarity: 100.000 Percent identity: 100.000

alignment\_block:  
US-09-357-675c-21 x AA246101 ..

Align seg 1/1 to: AA246101 from: 1 to: 1416

```

2  LeuGIypHeIIeThrArgProProHisArgPheLeuSerLeuLeuCypr 18
|||||
111 CCGGGCTTCATCACCAGCGGCTCTCTCACAGATTCTGTCCTCTGTC 160
18  oGILeuArGIleProGILeuSerValLeuCYsAlaGInProArGPro 35
|||||
161 TCGACTCCGGAATCCTCAACTCTCACTACTTGTGCTCAGCCAGGCCA 210
35  rGAlaMeTaIaIleSerSerSerSerCysGluLeuProLeuValAla 51
|||||
211 GAGCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 260
52  CysGluValThrSerThrProAspLysGlnGlnAsnPhelYThrCysAl 68
|||||
261 TCCACAGATACATCGACGCCAGACAAGCAACTTTAAACATGTGC 310
68  aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeu 85
|||||
311 TCGAGCTGGTTCAGAGGCTGCGCAGACTGGGCTCCCTGGCTTCTCC 360
85  roGluAlaPheAspHeIIeAlaArgAspProAlaGluThrLeuHisLeu 101
|||||
361 CCGAGGCAATTGACTTCATTCAGCGGACCTCTGACAGACGCTACACC 410
102  SerGIuProLeuGIyGIyLysLeuLeuGluGluTyrThrGIuLeuAla 118
|||||
411 TTGTAACCACTGGGTGGGAACTTTGGAGAAATACACCACGCTTCCG 460
118  gGIuCYsGIyLeuTrPleuSerLeuGIyGIyPheHisGIuArgGIyGln 135
|||||
461 GGAATGTGACTCTGGCTCTCTCTGGGTGGTTCATGAGCGGGCCAG 510
135  sPTrGIuGIuThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
|||||
511 ACTGGAGAGAGACTCAGAAATCTCAATTTGACCTGCTGTGAACACC 560
152  LysGIyAlaValAlaThrTyrArgLysThrHisLeuCysAspValGI 168
|||||
561 AAAGGGGCGAGTAGTGGCCACTTACAGAAAGACATCTGTGACGTAGA 610
168  uIleProGIyGlnGIyProMeCYsGIuSerAsnSerThrMeIProGIyP 185
|||||
611 GATTCACAGGCGAGGGGCTATGTGTGAAGCAACTTACATGCGTGGCC 660
185  roSerLeuGIuSerProValSerThrProAlaGIyLysIleGIyLeuAla 201
|||||
661 CCAGTCTTGAGTCACTCTGACACACCAGCGGCAAGATTGTGTAGCT 710
202  ValCYsTYrAspMeuArGpPheProGIuLeuSerLeuAlaLeuAlaGln 218
|||||
711 GCTCTGTATGACATGCGGTTCCCTGACACTCTCTGCGCAATGGCTCA 760
218  aGluGIuGIuLeuLeuThrTYrProSerAlaPheGIySerIleThrGIyP 235
|||||
761 TCGAGACAGAGATACCTATCCTTACCTTTTGGATCCATTAACAGGCC 810
235  roAlaHisTrpGIuValLeuLeuArgAlaArgAlaIleGIuThrGlnCYs 251
|||||
811 CAGCCCACTAGGAGGTGTCTGCGGGCCCTGTATCCAAACCCAGTGC 860
252  TYrValAlaAlaAlaGlnCysGIyArgHisHisGIuLysArgAlaSe 268
|||||
861 TATGTATGGCAGCAGACAGTGTGAGCGCCACATGAGAAAGAGCAAG 910

```

```

268 rTYrGIyHisSerMetValAlaAspProTrpGIyThrValAlaIaArgC 285
|||||
911 TATGCGCCACAGCATGATGGTAGACCCCTGGGGAACAGATGGTGGCCGT 960
285 ySSerGIuGIyProGIyLeuCYsLeuAlaArgIleAspLeuAsnTYrLeu 301
|||||
961 GCTGTGAGGGGCCAGGCTCTGCTTGGCCGAAATGACCTCAACTATCTG 1010
302  ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLe 318
|||||
1011 CGACAGTTGCGCCGACACCTGCTGTGTTCCAGCACCGAGGCCCTGACCT 1060
318  uTYrGIyAsnLeuGIyHisProLeuSer 327
|||||
1061 CTATGCAATCTGGGTCAACCCACTGTCT 1088

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAF16257

seq_documentation_block:
ID  AAF16257 standard; cDNA; 1382 BP.
XX
AC  AAF16257;
XX
DF  13-MAR-2001 (first entry)
XX
DE  Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX
KW  Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW  neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW  vulnerrary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW  anticancer; gene therapy; neutral; immune; reproductive; renal;
KW  gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW  wound; infectious disease; ss.
XX
OS  Homo sapiens.
XX
FN  W0200055174-A1.
XX
PD  21-SEP-2000.
XX
PF  08-MAR-2000; 2000WO-US05988.
XX
PR  12-MAR-1999; 99US-0124270.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
PA  (ROSE-) ROSEN C A.
XX
PI  Rosen CA, Ruben SM.
XX
DR  WPI: 2000-587513/55.
DR  P-PSDB: AAB57054.
XX
PT  Prostate cancer associated gene sequences, referred to as prostate
PT  cancer antigens, useful for treatment, prevention, and diagnosis of
PT  disorders such as prostate cancer -
XX
PS  Claim 1: Page 1124: 2338bp; English.
XX
CC  AAF15566 to AAF16505 encode the human prostate cancer associated
CC  proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC  The prostate cancer antigens can have neuroprotective, cytosolic,
CC  cardioactive, immunomodulatory, muscular, vulnerrary, gastrointestinal,
CC  nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC  and are used in gene therapy. The prostate cancer antigen
CC  polynucleotides may be used for detection of prostate cancer, chromosome
CC  identification, as chromosome markers, and for numerous other diagnostic
CC  or research purposes. The prostate cancer antigens may be used to treat
CC  disorders such as neural, immune, muscular, reproductive,
CC  gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC  disorders, wounds, and infectious diseases. AAF15506 to AAF16514 to
CC  AAB57303 represent sequences used in the exemplification of the present
CC  invention.
XX

```

50 Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

## alignment\_scores:

Quality: 1718.00 Length: 327  
Ratio: 5.270 Gaps: 0  
Percent Similarity: 99.694 Percent Identity: 99.694

## alignment\_block:

US-09-357-675c-21 x AAF16257 ..

Align seg 1/1 to: AAF16257 from: 1 to: 1382

```

1 MetLeuGlyPheIleThrArgProProHisThrPheLeuSerLeuLeu 17
|||||
77 ATGCTGGGCTTCATCCACGAGCTCCACAGATTCCTGCTCCCTGTGTG 126
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArg 34
|||||
127 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTGTGCTTCAGCCAGCC 176
34 roArGAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuValAla 50
|||||
177 CCAGAGCCATGGCTATGCTCTCTCTCTCTGCAACTGCCCTGTGGCT 226
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelyThrCy 67
|||||
227 GTGTGGCAGGTAAACATCGACCGCACAGACAGCAAGCACTTAAACATG 276
67 sAlaGlnLeuValArgGlnAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
277 TGTGACTGCTGTGAGAGCTGCGACAGCTGGGCTGCTGCTGCTTTC 326
84 euProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThrLeuHis 100
|||||
327 TGCTGAGGATTTGACTTCACTTCACGCGACCTGCAGAGACGCTACAC 376
101 LeuSerLeuProLeuGlyGlyLysLeuLeuGlnLurThrGlnLeuAl 117
|||||
377 CTGTGTGAACCATCTGGGTGGAACTTTGGAAATACACCCAGCTTGC 426
117 aArgGlyCysGlyLeuThrPheLeuSerLeuGlyGlyPheHisGlnArgLys 134
|||||
427 CAGGGAATGTGGACTTGGCTGTCTTGGGTGGTTCATAGCGTGGCC 476
134 InAspTrpGlnGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
|||||
477 AAGACTGGAGACAGACTCAGAAATCTACAAATGTCACGCTGCTGAC 526
151 SerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeuCysAspVa 167
527 AGCAAAAGGGCAGTAGTGGCCACTTACAGGAAGACATCTGTGTGAGCT 576
167 IGlulIleProGlyGlnGlyPrometCysGlnSerAsnSerThrMetProG 184
|||||
577 AGAGATTCCAGGCGAGGGGCT.ATGTGTGAAGCAACTCTTACCTGCTG 625
184 LyProSerLeuGlnSerProValSerThrProAlaGlyLysIleGlyLeu 200
|||||
626 GGCCCACTTGAAGTCACTGTCAGACACAGCAGCAGCAAGCAATTGGTCA 675
201 AlaValLysTyrAspMetArgPheProGlnLeuSerLeuAlaLeuAlaG 217
|||||
676 GCTGTCTGATGACATGCGGTTCCCTGAACTCTCTGCGATTGGCTCA 725
217 nAlaGlyAlaGlnIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
|||||
726 AGCTGGAGCAGAGTACTTACTTACCTTCCAGCTTTGGATCCATTACAG 775
234 LyProAlaHisTrpGlnValLeuLeuArgAlaArgAlaIleGlnThrGln 250
|||||
776 GCCCAGGCCCACTGGAGAGTGTGCTGGGGGCCGCTGCTATCGAAACCCAG 825

```

```

251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGlnLysArgAl 267
|||||
826 TGTATGTAGTGGCAGCAGCAGCACTGTGGACCCACCATGGAAGAGAGC 875
267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAla 284
|||||
876 AAGTTATGGCCACAGCATGCTGTAGACCCCTGGGAAACATGTGTGGGCC 925
284 rGysSerGlyGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
|||||
926 GCTCTCTGAGGGGCGCAGGCTCTGCTTGGCCCAATAGACCTCAACTAT 975
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
976 CTGGACAGTTGGCGCCGACACCTGCTGTTCAGACACCGAGGCTGA 1025
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
1026 CCTCTATGGCAATCTGGGTGACCCACTGTCT 1056
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAx30398
seq_documentation_block:
ID AAX30398 standard; DNA; 1203 BP.
XX
AC AAX30398;
XX
DT 14-MAY-1999 (first entry)
XX
DE DNA encoding a human secreted protein.
XX
KW Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS Homo sapiens.
XX
PN MO9907891-A1.
XX
PD 18-FEB-1999.
XX
PF 04-AUG-1998; 98MO-US16235.
XX
PR 19-AUG-1997; 97US-0056732.
PR 05-AUG-1997; 97US-0054798.
PR 05-AUG-1997; 97US-0054803.
PR 05-AUG-1997; 97US-0054804.
PR 05-AUG-1997; 97US-0054806.
PR 05-AUG-1997; 97US-0054807.
PR 05-AUG-1997; 97US-0054808.
PR 05-AUG-1997; 97US-0054809.
PR 05-AUG-1997; 97US-0055309.
PR 05-AUG-1997; 97US-0055310.
PR 05-AUG-1997; 97US-0055312.
PR 05-AUG-1997; 97US-0055386.
PR 05-AUG-1997; 97US-0055311.
PR 18-AUG-1997; 97US-0055970.
PR 18-AUG-1997; 97US-0055986.
PR 19-AUG-1997; 97US-0056365.
PR 19-AUG-1997; 97US-0056366.
PR 19-AUG-1997; 97US-0056557.
PR 19-AUG-1997; 97US-0056370.
PR 19-AUG-1997; 97US-0056371.
PR 19-AUG-1997; 97US-0056563.
PR 19-AUG-1997; 97US-0056731.
XX
PA (HDMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Ebner R, Ferlie AM, Greene JM, Janat F, Ni J;
Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;

```

XX WPI: 1999-167452/14.  
 DR P-PSDB: AAV10877.  
 XX New isolated human genes encoding secreted polypeptides - useful for  
 PT diagnosis and treatment of pathological diseases  
 PS  
 XX Claim 3; Page 265-266; 331pp: English.

XX The specification describes secreted proteins and their corresponding  
 CC polynucleotides which are useful for preventing, treating or ameliorating  
 CC medical conditions, e.g. by protein or gene therapy. Pathological  
 CC conditions can also be diagnosed by determining the amount of the  
 CC secreted polypeptides in a sample or by determining the presence of  
 CC mutations in the polynucleotides. Specific uses are described for each  
 CC of the products, based on which tissues they are most highly  
 CC expressed in, and include developing products for the diagnosis or  
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental  
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,  
 CC diseases of the immune system, autoimmune diseases, hepatic and renal  
 CC diseases, diabetes, inflammation, allergies, ischemic shock, Alzheimer's  
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,  
 CC prostate diseases, asthma, disorders involving osteoclasts such as  
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or  
 CC thymus, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.

SQ Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;

alignment\_scores:  
 Quality: 1067.00 Length: 207  
 Ratio: 5.205 Gaps: 0  
 Percent Similarity: 99.034 Percent Identity: 99.034

alignment\_block:  
 US-09-357-675C-21 x AAX30398 ..

Align seg 1/1 to: AAX30398 from: 1 to: 1203

121 GYLLEUTRPLEUSERLGGLYGLYPHEHISGLUARGLYSLNAPRPGI 137  
 ||||||||||||||||||||||||||||||||||||||||||||  
 3 GGCCTGTGGCTGCTTGGGTGGTTCCATGACGCGTGCCCAACATGGGA 52  
 ||||||||||||||||||||||||||||||||||||||||||||  
 137 uGlnTrGlnLysIleTyrAsnCySHisValLeuLeuAsnSerLysGly 154  
 ||||||||||||||||||||||||||||||||||||||||||||  
 53 GCAGACTTCAGAAATCTACAAATGTCACGTCGTCGAGAACAG.AAAGGG 101  
 ||||||||||||||||||||||||||||||||||||||||||||  
 154 lAValAlaIaIaTrTyrArgLysThHisLeuCySAspValGluIlePro 170  
 ||||||||||||||||||||||||||||||||||||||||||||  
 102 CAGTACTGGCCACTTACAGAGAACATCTGTGTGACGTAGAGATTCCA 151  
 ||||||||||||||||||||||||||||||||||||||||||||  
 171 GLYGLNGLYPrometCysGluSerAsnSerThMetProGlyProSerIle 187  
 ||||||||||||||||||||||||||||||||||||||||||||  
 152 GGGCAGAGGGCT.ATGTTGTAAGCAACTTACCATGCTGGCCCACTCT 200  
 ||||||||||||||||||||||||||||||||||||||||||||  
 187 uGluSerProValSerThProAlaGlyLysIleGlyLeuAlaValCyst 204  
 ||||||||||||||||||||||||||||||||||||||||||||  
 201 TGTGATCACTGTCAGACACACAGCAGCAAGATTGGTCTAGCTCTGCT 250  
 ||||||||||||||||||||||||||||||||||||||||||||  
 204 YrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220  
 ||||||||||||||||||||||||||||||||||||||||||||  
 251 ATGACATGGCGGTTCCCTGAACCTCTCTGTCATGTCATGAGCTGGAGCA 300  
 ||||||||||||||||||||||||||||||||||||||||||||  
 221 GluIleLeuThTyrProSerAlaPheGlySerIleThGlyProAlaH 237  
 ||||||||||||||||||||||||||||||||||||||||||||  
 301 GAATATCTTACCTATCCCTTGGATTCATTACACAGCCAGCCCA 350  
 ||||||||||||||||||||||||||||||||||||||||||||  
 237 STPRGluValLeuLeuArgAlaArgAlaIleGluTrnGlnCysTyrVal 254  
 ||||||||||||||||||||||||||||||||||||||||||||  
 351 CTGGAGAGTTCCTGCGGGCCCGTCTATGAAACCACTGATCTAGTAG 400  
 ||||||||||||||||||||||||||||||||||||||||||||  
 254 aAlaIaIaIaIaGlnCysGlyArgHisGluLysArgAlaSerTyrGly 270

||||||||||||||||||||||||||||||||||||||||||  
 401 TGGCAGCAGCAGCAGTGTGGAGCGCCACCATGAGAGAGAGCACTTATGCG 450  
 ||||||||||||||||||||||||||||||||||||||||||||  
 271 HisSerMetValValAspProTrpGlyThValValAlaArgCysSergI 287  
 ||||||||||||||||||||||||||||||||||||||||||||  
 451 CACAGCATGTGTGTAGACCCCTGGGGACAGTGTGTGGCCCTGCTCTGA 500  
 ||||||||||||||||||||||||||||||||||||||||||||  
 287 uGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnL 304  
 ||||||||||||||||||||||||||||||||||||||||||||  
 501 GGGGCCAGGCGCTTGCCTTGGCCGAATAGACCTCACTATCTCGACAGT 550  
 ||||||||||||||||||||||||||||||||||||||||||||  
 304 euArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 320  
 ||||||||||||||||||||||||||||||||||||||||||||  
 551 TGGCGCGACACCTGCGCTGTCTCCAGACCGCAGCGCTGACCTATATGCG 600  
 ||||||||||||||||||||||||||||||||||||||||||||  
 321 AsnLeuGlyHisProLeuSer 327  
 ||||||||||||||||||||||||||||||||||||||||||||  
 601 AATCTGGGTACCCACTGTCT 621

seq\_name: /SIDS1/gcgsdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: AAD25458

seq\_documentation\_block:

ID AAD25458 standard; cDNA; 1214 BP.

XX AAD25458;

DT 26-MAR-2002 (first entry)

DE Xenopus laevis Nt1l cDNA.

KW Frog; Nt1l; cytosolic; neuroprotective; cellular pathway; therapy;

KW apoptosis; proliferative disorder; degenerative disease; ss.

XX Xenopus laevis.

FT Key Location/Qualifiers

FT CDS 180..1046

FT /tag= a

FT /product= "Frog Nt1l protein"

PN W0200187958-A2.

PD 22-NOV-2001.

PF 15-MAY-2001; 2001WO-US15664.

PR 16-MAY-2000; 2000US-204713P.

PA (UIJE-) UNIV JEFFERSON THOMAS.

PI Croce C, Brenner C, Bekarski Y;

DR WPI: 2002-082984/11.

XX P-PSDB: AAE15792.

PT Isolated cDNA encoding human, mouse, frog and yeast Nt2 proteins,

PT useful to find molecules that mimic or antagonize Fhit interaction for

PT the treatment of proliferative or degenerative diseases -

PS Claim 21; Page 60; 61pp: English.

XX The invention relates to isolated nucleic acids comprising a fully

CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and

CC mouse Nt2 proteins. Nt1 and Fhit proteins are encoded as fusion

CC proteins in invertebrates and as separate polypeptides in vertebrates.

CC Nt1 and Fhit interact physically and functions in same cellular pathways.

CC Molecules which bind Nt2 and mimic or antagonize Fhit interaction are

CC used to treat diseases in which activity of Nt2 protein is altered in

CC a mammal. Fhit mimics induce apoptosis and are particularly useful to

CC treat proliferative disorders, whilst Fhit antagonists promote cell

CC proliferation and are particularly useful to treat degenerative disease.

CC The present sequence is frog Nt1l cDNA.



Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;

# alignment\_scores:

Quality: 968.50 Length: 307  
Ratio: 3.905 Gaps: 1  
Percent Similarity: 80.782 Percent Identity: 57.980

## alignment\_block:

US-09-357-675c-21 x AAD25458 ..

Align seg 1/1 to: AAD25458 from: 1 to: 1214

```

16 LeucylProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
   |||||||.....:|||||.....
144 CTGTGTCGGATATAGGGGTTAGAAAGCTGTGATGCAATG..... 182
32 cArgProArgAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuV 49
   .....:|||||.....GCTGGTCCCAACAGCCCTGTA 204
183 .....
49 aAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelLys 65
   :|||||.....:|||||.....
205 TTGCGGTGGCCAGATGACTTCAACCTCTGATTAAGAGAAAGAAATTCGCC 254
66 ThrCysAlaGlnLeuValArgGlnAlaAlaArgLeuGlnAlaCysLeuAl 82
   |||||||.....:|||||.....
255 ACGTTTCGGCGGTATCCGGGAGCTCCGGGCGTCCGCTTCATGTCATGTC 304
82 aPheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThrL 99
   :|||||.....:|||||.....
305 GTTTCGCCGAAGCCTTTGACTATATCGGGGCGACGATTAAGAGAGAGCC 354
99 eunHisLeuSerGlnProLeuGlnLysLeuGlnGlnLysThrGln 115
   || |||||.....:|||||.....
355 TGAGCTGTGGCTGAGTCTCTACATGGGGACACCATTCACGTTACACCCAA 404
116 LeuAlaArgGlnCysGlnLeuThrLeuSerLeuGlnGlnPheHisGlnAr 132
   |||||||.....:|||||.....
405 CTCGCCAGGAGTGTGGCTGTGCTTCCTGGGGGATTCATGAGAA 454
132 gGlyGlnAspTrpGlnGlnThrGlnLysIleTyrAsnCysHisValLeuL 149
   :||| :|||.....:|||||.....
455 AGGACCCCAACTGGGACAGGACGACCAACGATTCATTCACAGGCGTGG 504
149 euaAsnSerLysGlyAlaValAlaAlaThrTyrArgLysThrHisLeuL 165
   :||| :|||.....:|||||.....
505 TGGACACACAGGCGACATAGTACGTGACCGCAAGGCTCACCGTGT 554
166 AspValGlnLeuProGlnGlnLysPrometCysGlnSerAsnSerThrMe 182
   |||||||.....:|||||.....
555 GACGTAGCTTGCAGAAATGAGTGTCACTCAAGAGAGCTTCACCCCT 604
182 tProGlyProSerLeuGlnSerProValSerThrProAlaGlnLysIleG 199
   :||||| :|||.....:|||||.....
605 CCCGGGACAGACGCTTATTCGCCCATCACTTCTCCACAGAAAGATTCG 654
199 lylLeuAlaValCysTyrAspMetArgPheProGlnLeuSerLeuAlaLeu 215
   |||||||.....:|||||.....
655 GCCTGGGGGTGTGTACGACCTCCGCTCCCAAGATTCCTCTGGCTGTG 704
216 AlaGlnAlaGlyAlaGlnLeuLeuThrTyrProSerAlaPheGlnSerI 232
   ||||||| :|||.....:|||||.....
705 GCCCCAACAGGAGCAGACCTCTCACTTACCTTCTGCTTCACCCCTCAG 754
232 eThrGlyProAlaHisTrpGlnValLeuLeuArgAlaArgAlaIleGlnT 249
   ||||||| :|||.....:|||||.....
755 TACGTGTGTGCACATGTGGAGGTCTTCTGAGAGCCGCTGCCAATGAA 804
249 hrGlnCysTyrValValAlaAlaGlnCysGlyArgHisHisGlnLys 265
   ||||||| :|||.....:|||||.....
805 CCCAGTGTCTAGTAGTGTGACGCGACAGACAGACAGACACAAATGAGAG 854

```

```

266 ArgAlaSerTyrGlnHisSerMetValValAspProTrpGlnThrValVa 282
   |||||||.....:|||||.....
855 AGGACGTCTATGTGTACGCGTATGTGTGTAGACCCGTGGGGGTGTCTAT 904
282 lAlaArgCysSerGlnGlnGlyProGlnCysLysLeuAlaArgIleAspLeuA 299
   :||||| :|||.....:|||||.....
905 TGGCCAAATGCCAGAGAGACAGAAATGTATTATGCTGATGACATTC 954
299 snTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArg 315
   |||||||.....:|||||.....
955 CCTACATGAGACCGTGTGAGCGGAGACATGCCGTGTGGAGCCGACGAG 1004
316 ProAspLeuTyrGlnLysLeu 322
   |||||||.....
1005 ACTGATCTGTATGGGAAATTC 1025

```

seq\_name: /sids1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.ABL12225

seq\_documentation\_block:

ID ABL12225 standard; cDNA; 1495 BP.

AC ABL12225;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-P-SDB; ABB68122.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 31157; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA

CC sequences (AB101840-AB116175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;

## alignment\_scores:

Quality: 710.50 Length: 294  
Ratio: 3.259 Gaps: 3  
Percent Similarity: 74.150 Percent Identity: 48.639



```

1183 GGATCAAAAGTGGACCATTTGCTGTGGTGCAGATGCCCTCTACACGAGACA 1232
      :      :      :      :      :      :      :      :      :
60 ysgInglInAsnPhelysthrcysAlaGluLeuValArgGluAlaAlaArg 76
      :      :      :      :      :      :      :      :      :
1233 AGCGCGCTATCTTTACCAAGTGATGAGCTAGAGCTAGTAGAGCCCAAGTCA 1282
      :      :      :      :      :      :      :      :      :
77 LeuGlyAlaCysLeuAlaPhelLeuProGluAlaAlaPhaAspPhelAlaAr 93
      :      :      :      :      :      :      :      :      :
1283 CAGAAAGCCTGCATGCTCTTTCCTGAGTGCTGAGCTTTGCTGGGCGA 1332
      :      :      :      :      :      :      :      :      :
93 gaspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeu 110
      :      :      :      :      :      :      :      :      :
1333 GAGCCGCGACCAAACTATGAGCTCTCCAGCGCTTGACGAGCGGAGTTAA 1382
      :      :      :      :      :      :      :      :      :
110 euGluGluThrThrGluLeuAlaArgGluCysGlyLeuThrLeuSerLeu 126
      :      :      :      :      :      :      :      :      :
1383 TGGCCGAGTACCGGGAATTGGCAAGTGCACAAATTTGATTTCCCTG 1432
      :      :      :      :      :      :      :      :      :
127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnIleTy 143
      :      :      :      :      :      :      :      :      :
1433 GGTGGGCTGCAGAGCGGAACGAT.....CAAAAATCTT 1467
      :      :      :      :      :      :      :      :      :
143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaThrTyra 160
      :      :      :      :      :      :      :      :      :
1468 CAACGCTCATGTTTGTCTCAGCAGAAAGGAGGAGACTAGCAGCATACA 1517
      :      :      :      :      :      :      :      :      :
160 rGlyThrHisLeuCysAspValGluIleProGlyGlnGlyPrometCys 176
      :      :      :      :      :      :      :      :      :
1518 GAAAGCTGCACATGTTGATGTACGACTAAA...GAGGTCCGCTAACGC 1564
      :      :      :      :      :      :      :      :      :
177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProAlaSerh 193
      :      :      :      :      :      :      :      :      :
1565 GATCTGATGATAGTTAGCGCGGATCTGCTGAGCGCCCATGAGAC 1614
      :      :      :      :      :      :      :      :      :
193 rProAlaGlyLysIleGlyLeuAlaValCysTyraAspMetArgPheProG 210
      :      :      :      :      :      :      :      :      :
1615 TCCAGTTGCCAGATAGGGCTTCAGATTGCTACGACCTGCGTTGGCG 1664
      :      :      :      :      :      :      :      :      :
210 LuLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyPro 226
      :      :      :      :      :      :      :      :      :
1665 ACCCGCGCTGCTGCTCAGGAAGCTGGGCCAATTTGTTAAACATACCA 1714
      :      :      :      :      :      :      :      :      :
227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
      :      :      :      :      :      :      :      :      :
1715 TCCGCAATTCATATACGAACCGGTAAAGCGCACTGGGAAATCCTCTGCG 1764
      :      :      :      :      :      :      :      :      :
243 gAlaArgAlaIleGluThrGlnCysTyraValAlaAlaAlaGlnCysG 260
      :      :      :      :      :      :      :      :      :
1765 GGGCAGAGCATAGAGACTCAATGCTTTGGTGGCTGCGCGCTCAAGTAC 1814
      :      :      :      :      :      :      :      :      :
260 LysArgHisIleGluLysArgAlaSerTyGlyHisSerMetValAlaAsp 276
      :      :      :      :      :      :      :      :      :
1815 GTTGGCACAACGACGACAGACAGATGGGGCCACAGCATGATCGTTAGC 1864
      :      :      :      :      :      :      :      :      :
277 ProTrpGlyThrValAlaAlaArgCysSerGluGlyPro...GlyLeu 292
      :      :      :      :      :      :      :      :      :
1865 CCTGGGGAACGTAAGCTGCTGAGCTCAGAGCAGAGAGCTGATATAGC 1914
      :      :      :      :      :      :      :      :      :
292 sLeuAlaArgGlyLeuAspLeuAsnTrpLeuArgGlnLeuArgArgHisLeu 309
      :      :      :      :      :      :      :      :      :
1915 CAGCGCGAGAGTGAGACTTTCCCTGCTCAATCTCTGTATCAGACCATGC 1964
      :      :      :      :      :      :      :      :      :
309 roValPheGlnHisArgArgProAspLeuTy 319
      :      :      :      :      :      :      :      :      :
1965 CCGCTTCGAAACATCGTCAAAACGACATCATC 1996
seq_name: /SIDSL/gcdata/hold-geneseq/genesegn-emb1/NA2001B.DAT:ABL20222
seq_documentation_block:
ID   ABL20222 standard; DNA; 5692 BP.
XX
AC   ABL20222;

```

```

XX
DE 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12139; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WFO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;

```

```

alignment_scores:
  Quality: 710.50      Length: 294
  Ratio: 3.259        Gaps: 3
  Percent Similarity: 74.150  Percent Identity: 48.639

```

```

alignment_block:
US-09-357-675C-21 x ABL20222

```

```

Align seg 1/1 to: ABL20222 from: 1 to: 5692

```

```

27 ValLeuCysAlaGlnProArgProArgAlaAlaAlaIleSerSerSer 43
   |||:||||:  :  :  :  ||| |||:||||:  :  :
2792 GTGATCGCATTCATCAGCAGCTACGCGAGATGCTCCCAAAAGAGAA 2841
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAsp 60
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
2842 GGATCAAAAGTGCAGCAATGCTGTGGTGCAGATGCGCTTACACAGGACA 2891
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
60 ysgInglInAsnPhelysthrcysAlaGluLeuValArgGluAlaAlaArg 76
   |||:||||:  :  :  :  ||| |||:||||:  :  :
2892 AGCGGCTATCTTTACCAAGTGATGAGCTAGAGCTAGAGCCCAAGTCA 2941
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
77 LeuGlyAlaCysLeuAlaPhelLeuProGluAlaAlaPhaAspPhelAlaAr 93
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
2942 CAGAAAGCCTGCATGCTTTCCTGAGTGCAGTGTGATTTGTTGGGCGA 2991
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
93 gaspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeu 110

```

```

2992 GAGCGCGAAGCCAAACTATTATTGAGCTTCGCGAGCGGCTTGGACGCGCAAGTTAA 304.1
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
110 euglucgluTyrThrGlnLeuAlaArgLysGlyLeuTrpLeuSerLeu 126
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3042 TGGCGCAGTACCGGGAATTTGGCGAAGTCACAAATAATTTGGATTTCCTG 309.1
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
127 GLyGlyPheHisGluArgGlyGlnAspTrpGlnGlnThrGlnLysIleTyr 143
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3092 GGTGGCGTCCACAGACCGGAACCAT.....CAAAAATTCCT 312.6
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
143 rAnScyHisValLeuLeuAsnSerLysGlyValAlaValAlaThrTyrA 160
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3127 CAACGCTCACTGTTTGGCTTCACGAGAAAGGGGAACTACACACAGTATACA 317.6
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
160 rglTyrHisLeuCysAspValGluLeuProGlnGlnLysProMetCys 176
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3177 GAAGCGTCACATGTTTGATGTATGACTATAA...GAGGTTTCGGCTACGC 322.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
177 GluSerAsnSerThrMetCpGlyProSerLeuGlnSerProValSerTh 193
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3224 GATATGAGTACAGTATACCCCGGGATACAGCCCTTGAGCCGCCAGAGACAC 327.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3274 TCCAGTGGCCAGATAGGGGCTTCAGATTGGCTACACCCCTGGCTTTGGCTG 332.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
210 LueSerLeuAlaLeuAlaGlnAlaGlyValAlaGlnIleLeuThrTyrPro 226
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3324 AGCGCGCGGTGCTGCTCAGAGAACGCGGGTCCCAATTTGTTAACTATCCA 337.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3374 TCCGCATTCACATACGCAACCGGTAAGCGCAGCTGGAAATCTCTCGCG 342.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
243 GAlaArgAlaIleGlyThrGlnCysTyrValValAlaAlaAlaGlnCysG 260
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3424 GGCACAGCGCAATGAGACTCAATGCTTTGGTGGCTCGCTCAGAGTAC 347.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
260 LVArGHsHisGluLysArgAlaSerTyrGlnLysSerMetValValAlaAs 276
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3474 GTTGGCAACAACGAGAGGACAGAGTATGGGGCCACAGCATGATGCTTAAGC 352.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
277 ProTrpGlyThrValValAlaAlaArgCysSerGlnGlyPro...GlyLeuCy 292
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3524 CCTGGGGAAACGTACTGGCTACAGCGAGCGAGAGAGCTGTATATAGG 357.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
292 sLeuAlaArgIleAspLeuAsnTyrTLeuArgGlnLeuArgArgHisLeuP 309
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3574 CAGGCGCGAGGTGAGACCTTTCCGTCCTTCAATTCCTGATACAGACATGC 362.3
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
309 rAlaValPheGlnHisArgArgProAspLeuTyr 319
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3624 CCTGCTTGAAACATGCTGAAACGACATCTAC 365.5
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF84214

seq\_documentation\_block:

ID AAF84214 standard; cDNA; 1001 BP.

AC AAF84214;

DT 12-JUN-2001 (first entry)

Human hnit3-ase coding sequence.

KW Human; hnit3-ase; Cushings adrenal gland tumour;

nitrite hydrolytic enzyme; nitrilase; ss

05 Homo sapiens

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

F <sup>IT</sup>	CDS	99...929
-----------------	-----	----------

```
FT      /tag= a
FT      /product= "Human hnit3-ase"
```

PN CN1277998-A.

PD 27-DEC-2000.

PF 30-MAY-2000; 2000CN-0116221

PR 30-MAY-2000; 2000CN-0116221

PA (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.

PI Xu X, Qian B, Zhang X;

DR WPI; 2001-245678/26

[illegible]

PS Claim 1; page 16-17; 20pp; Chinese.

CC The present invention relates to human hnitf-ase (nitirfase) protein,  
CC which is expressed in Cushing's adrenal tumour, and its coding  
CC sequence (AA084214 and AAB80984). The present invention also relates to a  
CC preparation method of said protein and nucleic acid sequence, and a  
CC method of detecting human hnitf-ase nucleic acid sequence and polypeptide  
CC in sample.

Sequence 1001 BP; 264 A; 235 C; 245 G; 257 T; 0 other;

alignment\_scores:

Quality:	495.00	Length:	321
----------	--------	---------	-----

Percent Similarity: 64.174    Percent Identity: 34.891

```
alignment_block:
```

Align seg 1/1 to: AAF84214 from: 1 to: 1001

26 SerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSer 42

11 GCATTATGGCCGGGAGGCCGACGCCCGCGGTCCTGATCTCCAGGCG 60

42 rSercysGIuLeu.ProLeuValAlaValcysGlnValIhrSer..... 5b

61 CTTCAGTCCGCGCCGACAGGTGGTGTCTTGTCTTCACAGAGTCAATGACCTCTTTC

5 / .....inrproaspplyscnglnnashn 04  
.....

111 CGCTGGCCCTCATCCAGCTTCAGATTCTCCATCATTCCAGTAAAGG 100

64 elysinrcysalagluLeuValIrgluAlaAlaIrgLeuGlyAlaCysL 81

161 CACTCGGCTTGAGCTTCATCCGGAGGCCAGCAACCCAGGAGCCATAA ZI

81 euaIaPheLEUProGLuaIaPheaspPheILeaIaArgaspPheOlaIaGlu 3'

211 TAGTTTCTTTGGCCGGAATGCTTT..... 233

98 ThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyr.. 113

234 .....ATTTCCTCCATATGGAGCGAATAATTTCCTGAAATAGC 211

114 .....TINGLINLEUADALDAYS 115  
.....| | | | |

2 / 2 AGAGAAATTTCTTGATTCACACAGAAGCTTTCTTGAACTAGCAAGG 321

119 lucysglyleutirpleussetleuglyglyphenlsgualgolglnasr 120  
|||||.....|||

322 AATGCAGCATATACTTCATTGGAGGCTCTATCCCTGAA..... 323

```

136 TrrgluInthrGlnLysIleTyrAsnCysHisValleuLeuAsnSerly 152
    |||..... |||..... |||..... |||..... |||..... |||.....
360 ...GAGGATGCTGGGAAATATATATACACCTGCTGTGTGGCTGA 406
152 sGlyAlaValAlaIatThrTyrArgLysThrHisLeuCysAspValGlu 169
    |||..... |||..... |||..... |||..... |||..... |||.....
407 TGGACCTTACTAGCAAGATATAGAAAGATCCATCTTTGACACTTGATG 456
169 lrrpGlyGlnGlyPrometCysGluSerAsnSerThrMetProGlyPro 185
    |||..... |||..... |||..... |||..... |||..... |||.....
457 TTCCCTGGAAAAATTCATTTCAAGAAATCTAAACATTCAGAGCCGGGAT 506
186 SerLeuIleuSerProValSerThrProAlaGlyLysIleGlyLeuAlaVa 202
    |||..... |||..... |||..... |||..... |||..... |||.....
507 AGTTTC...TCCACATTTGATACCTCTTACTGTCGAGAGTGGGTCAT 553
202 lCysTyrAspMetArgPheProGlyLeuSerLeuAlaLeuAlaGlnAlaG 219
    |||..... |||..... |||..... |||..... |||..... |||.....
554 CTGCTAGACATGGGCTTTGCAGAGCTTGCAACAATCTAGCCACAGAGAG 603
219 lYalagIuIleLeuThrTyrProSerAlaPheGlySerIleThrGlyPro 235
    |||..... |||..... |||..... |||..... |||..... |||.....
604 GCTGCCAGCTGTGGTATATCCAGAGACCTTTTATCTGACCACCTGACCA 653
236 AlaHisTrpIuValLeuLeuAlaArgAlaIleGluThrGlnCysTy 252
    |||..... |||..... |||..... |||..... |||..... |||.....
654 GCCCATTTGGAGATCTTCACGCAAGCCGGCTGTGATATATCAAGGTGA 703
252 rValAlaAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSer 269
    |||..... |||..... |||..... |||..... |||..... |||.....
704 TGTGGCCACAGCCTCTCTGCTCC.....CGGATGACAAACCTCTCT 744
269 Yr.....GlyHisSerMetValAlaSprProTrpGlyThrValAla 282
    |||..... |||..... |||..... |||..... |||..... |||.....
745 ATGTTCCTGGGGACACGACACCTGGTGAACCTTGGGGAGAGTCTTA 794
283 AlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAs 299
    |||..... |||..... |||..... |||..... |||..... |||.....
795 GCCAAAGCTGGCACAGAAAGAACATCTGTATTCAGACATAGACCTGAA 844
299 nTyrLeuArgGlnLeuAlaArgArgHisLeuProValPheGlnHisArgAr 316
    |||..... |||..... |||..... |||..... |||..... |||.....
845 GAGGCTGGCGGAATATAGCCACGAAATCCCGCTTTTATAGACAGAGCAT 894
316 roAspLeuTyr 319
    |||..... |||..... |||..... |||..... |||..... |||.....
895 CAGACCTCTAT 905
seq_name: /stddsi/gcgdata/hold-geneseq/geneseg-emb1/NA2002.DAT.AAD25456
seq_documentation_block:
ID AAD25456 standard; cDNA; 1359 BP.
XX
AC AAD25456;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human Nlt2 cDNA.
XX
KW Human; Nlt2; cytosolic; neuroprotective; cellular pathway; therapy;
KW apoptosis; proliferative disorder; degenerative disease; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 22..852
FT /tag= a
FT /product= "Human Nlt2 protein"
XX
PN WO200187958-A2.
XX
PD 22-NOV-2001.

```

```

XX
PF 15-MAY-2001: 2001MO-US15664.
XX
PR 16-MAY-2000; 2000US-204713P.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Croce C, Brenner C, Pekarski Y;
XX
DR WPI; 2002-082984/11.
XX
DR P-PSDB; AAE15790.
XX
PT Isolated cDNA encoding human, mouse, frog and yeast Nlt2 proteins,
PT useful to find molecules that mimic or antagonise Fhit interaction for
PT the treatment of proliferative or degenerative diseases
XX
PS Claim 3; Page 59; 61pp; English.
XX
CC The invention relates to isolated nucleic acids comprising a fully
CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and
CC mouse Nlt2 proteins. Nlt and Fhit proteins are encoded as fusion
CC proteins in invertebrates and as separate polypeptides in vertebrates.
CC Nlt and Fhit interact physically and functions in same cellular pathways.
CC Molecules which bind Nlt2 and mimic or antagonise Fhit interaction are
CC used to treat diseases in which activity of Nlt2 protein is altered in
CC a mammal. Fhit mimics induce apoptosis and are particularly useful to
CC treat proliferative disorders, whilst Fhit antagonists promote cell
CC proliferation and are particularly useful to treat degenerative disease.
CC The present sequence is human Nlt2 cDNA.
XX
SQ Sequence 1359 BP; 388 A; 284 C; 312 G; 374 T; 1 other;

```

```

alignment_scores:
Quality: 485.50 Length: 286
Ratio: 2.542 Gaps: 6
Percent Similarity: 66.783 Percent Identity: 36.713

```

```

alignment_block:
US-09-357-675C-21 x AAD25456 ..

```

```

Align seg 1/1 to: AAD25456 from: 1 to: 1359

```

```

49 ValAlaValAlaGlnValThrSerThrProAspLysGlnGlnAsnPhely 65
    |||..... |||..... |||..... |||..... |||..... |||.....
37 TTGGCCCTCATCCAGCTTCAGATTTCTTCATCAATCAATCAGATACGTCAC 86
65 sThCysAlaGlnLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuA 82
    |||..... |||..... |||..... |||..... |||..... |||.....
87 TCGGCGCTGTAGCTTCATCCGGAGCGACGCAAGCAGAGCAACAAAATAG 136
82 lApheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThr 98
    |||..... |||..... |||..... |||..... |||..... |||.....
137 TTCTTTCCCGAATGCTTT..... 156
99 LeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyr.... 113
    |||..... |||..... |||..... |||..... |||..... |||.....
157 .....AATCTCCATATGAGCGAAATATTTTCTCGATATATGACA 197
114 .....ThrGlnLeuAlaArgGlnC 120
    |||..... |||..... |||..... |||..... |||..... |||.....
198 GAAATTCCTGCTGATCCACACAGAGCTTTCTGAAATACCAAGAAAT 247
120 ySgIlyLeuTrpLeuSerLeuGlyLysPheHisGluArgGlyGlnAspTrp 136
    |||..... |||..... |||..... |||..... |||..... |||.....
248 GCAGCATATATCTCATTTGAGGCTCTATCCCTGAA..... 282
137 GluGlnThrGlnLysIleTyrAsnCysHisValleuLeuAsnSerlyGcl 153
    |||..... |||..... |||..... |||..... |||..... |||.....
283 GAGGATGCTGGGAAATATATATACACCTGCTGTGTGGGCGCTGATGG 332
153 yAlaValAlaIatThrTyrArgLysThrHisLeuCysAspValGluIleP 170
    |||..... |||..... |||..... |||..... |||..... |||.....

```



PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other *Neisseria* infections, for example, *N. gonorrhoea* -  
 PS Claim 7; Page 547-567; 1760pp; English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25683 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC *Neisseria* bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus B*; against all serotypes;  
 CC and/or against all pathogenic *Neisseria*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX  
 SO Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

alignment\_scores:      Quality: 404.00      Length: 307  
                          Ratio: 2.138              Gaps: 8  
 Percent Similarity: 61.564      Percent Identity: 35.179

alignment\_block:

US-09-357-675C-21 x AAA81479/rev ..

Align seg 1/1 to reverse of: AAA81479 from: 1 to: 69936

```

16 LeucylProglyLeuArgIleProGlnLeuSerValLeucylAlaGlnPr 32
   |||||
27227 TTGTGTAAA.....CGTTTTTCGCAAGTACCGTCATGTAGTACCTATCG 27184
   |||||
32 ORF.....ProArgIleMetAlaIleSerSerSerSerSerCysGlnLeuPr 47
   |||||
27183 GCGGAATATGCGCGCAAGAGTCCGCGCAAGAAAGAAAGAAATGACAAATA 27134
   |||||
27133 TCACAGTTGCCCGCGCGTGCAGATGCTGCGCGCGTGCAGCAACAC 27084
   |||||
64 PheIysThrCysAlaGluLeuValArgIleAlaAlaArgLeuGlyAlaCy 80
   ::::
27083 GTCGCCCGCATGAAGCGCTGTGCACGCGCGCGCGAGCGGTGCGGA 27034
   ::::
80 sleuAlaPheIeuProGlnAlaIleAspPheIleAlaArgAspProIaG 97
   ::::
27033 TTGGTGCTGTCTCCGCAATATTGGGTCTATGTGCGCAACATATCCG 26984
   ::::
97 IuThrLeuHisLeuSerGluProLeu...GlyGlyIysLeuLeuGluGlu 112
   ::|||
26983 ACAAACTCGCGCTTCCGAGCGCTTGGCGCGCGGACCTTTCACAGCGCA 26934
   ::::
113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 128
   ::::
26933 TTGAGCGAAACGCGCAAGATGCGCGGTGTGCTG...TTCCGCGCGGAC 26887

```

```

|||||
409 GGGAAATATTCGTTCAAGATCTAAACATTCAGCCCTGTGATGTT 458
187 uGUSerProValSerThrProAlaGlyIleGlyLeuAlaValCyst 204
   ::|||
459 C...TCCACATTTGATACGCTTACCTGCAAAATGGCGCTGGCATCTGCT 505
   |||||
204 YrAspMetArgPheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
   |||||
506 ATGATATGCGCTTCGCGAGCTTGCAAAATCATGCAAAAGAGCTGC 555
221 GluIleLeuThrTyProSerAlaPheGlySerIleThrGlyProAlaH 237
   ::::
556 CACCTCTGTGTTATCTGTGAGCTTCAATCTGACACAGACAGACGCCA 605
237 strPGluValLeuLeuAlaArgAlaIleGluThrGlnCysTyValY 254
   |||||
606 CTGGGAGCTGCTTCAGCAGCCCGGCTGTGATATACAGTGTATGTG 655
254 aAlaAlaAlaGlnCysGlyArgHisHisGluIysArgAlaSerTyr... 269
   ::::
656 CTACACCTCTCTGCT.....CGGATGACAAAGCTGTATGTG 696
270 .....GlyHisSerMetValValaAspProTrpGlyThrValAlaAla 284
   |||||
697 GCGTGGGACACAGACACTGTTGTGATCTTGGGCGAGGCTCTAACCA 746
284 gCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrL 301
   ::::
747 AGCTGGCAGCGAGAAACAATCTGTACTGACATAGACCTGAGAGAGC 796
301 euArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAsp 317
   |||
797 TGGCTGAATTCGGCAGCAATCCCATTTTAAACAGAAACGACGAGAC 846
318 LeuTyr 319
   |||||
847 CTCTAT 852
seq_name: /SID1/gcdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAA81479
seq_documentation_block:
ID AAA81479 standard; DNA; 69936 BP.
AC AAA81479;
XX
XX 04-DEC-2000 (first entry)
DT
XX
XX N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.
DE
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
XX Neisseria meningitidis.
OS
XX WO200022430-A2.
PN
XX 20-APR-2000.
PD
XX
XX 08-OCT-1999; 99WO-US23573.
PF
XX
XX 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
PA
XX (CHIR ) CHIRON CORP.
XX
XX Frizer CM, Hickey E, Peterson J, Tetelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarcelli M, Scarlato V;
PI Rappuoli R, Pizzi M;
XX
XX WPI; 2000-318079/27.
XX

```

```

129 .....PheHisGluArgGlyGlnAspTrpGlnGlnThrGlnL 141
26886 TGTGGCGGTGCAGAACCTGCGAGGGGGT.....A 26858
141 yslEThyrAsnCysHisValLeuLeuAsnSerTylGlyAlaValAla 157
||||| ||| :||| :||| :||| :||| :||| :||| :|||
26857 AAGTGTATGAATACGCTGTGGGTATCGGACGAGCGGGGTAAAGACGGG 26808
158 ThrTyrAlaGlyTyrHisLeuCysAspValGluLeuProGlyGlnGlyPr 174
||||| ||| ||||| ||| :||| |||
26807 CTGTACCAACAAATATCCACTC.....TTCCGTTTTCGGGTTGGGCGGA 26764
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
||||| :||| :||| :||| :||| :||| :||| :||| :|||
26765 ACGTATGCGCGAAGCCGATACCATCCGCGCGGGGGGATGTCGGCAGT 26714
191 alSerThrProAlaGlyLysIleGlyLeuAlaValAlaCysTyrAspMetArg 207
||||| :||| :||| :||| :||| :||| :||| :||| :|||
26713 TGTGCGGAGAAAGCGCTGGCGGGGCGGGGCGCATTTGTATGATGTCCG 26664
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluLeuLeuTh 224
||||| ||| :||| ||| :||| :|||
26663 TTTCCCGAA.....TTTTCGACGCGCAGTGGCGTTTGACATGTGAT 26620
224 TrpTrpSerSerAlaPheGlySerIleThrGlyProAlaHisTrpLeuAl 241
||||| :||| :||| :||| :||| :||| :||| :||| :|||
26619 GCTGCCCTCGGTTTACGACACGACGCGGCGAAGCGCATTTGGAGCTGC 26570
241 euLeuAlaGAlaArgAlaIleGluThrGlnCysTyrAlaValAlaAlaAla 257
||||| :||| :||| :||| :||| :||| :||| :||| :|||
26569 TCTCTGCGCGCGGTCCCTCGAAAACCAATTTACGTGCTGGCGGGCGCA 26520
256 GlnCysGlyAlaArgHisIleGluAspArgAlaSerTyrGlnHisSerMet 274
||||| :||| :||| :||| :||| :||| :||| :||| :|||
26519 CAGGGCGGGTTTGCACGAAACGACGCGGCGACGTTCCGACACACATGAT 26470
274 ValAspProTrpGlyThrValAlaAlaArgCysSerGluGlyProGlyL 291
||||| :||| :||| :||| :||| :||| :||| :||| :|||
26469 TGTCCATCCGTGGGCGACGTGTGGACGATTTGCCGAGGCGGAAAGCG 26420
291 euCysLeuAlaArgIleAspLeuAsnTyrLeuAlaArgGlnLeuAlaArgHis 307
||||| :||| :||| :||| :||| :||| :||| :||| :|||
26419 TTGTTACGCGACACATGATGCCAACCCGCTGAACACAGCGTCGCAACCG 26370
308 LeuProValPheGlnHisArg 314
||||| :||| :||| :||| :||| :||| :||| :||| :|||
26369 CTGCCGCGCTTGAATACCG 26349

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAF21607
seq_documentation_block:
ID AAF21607 standard; DNA; 349980 BP.
AC AAF21607;
XX
XX 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence seq ID NO:108.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN MO200006791-Al.
XX
XX 09-NOV-2000.
XX
XX 08-MAR-2000; 2000MO-US05928.
XX
XX 30-APR-1999; 99US-0132068.
XX
XX 08-OCT-1999; 99MO-US23573.
XX

```

```

PR   28-FEB-2000; 2000GB-0004695.
XX
XX   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
XX   Pizza M, Hickey E, Peterson J, Tettein H, Venter JC, Masignani V;
PI   Galicotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappunoli R;
PI   Fraser CM, Grandi G;
XX
XX   WPI, 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -
XX
XX   Claim 7, Appendix A; 692pp; English.
XX
XX   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own It was split into 8
CC   sequences which overlap each other at the beginning and end of each
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21568 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   the present invention. The NMB genome and fragments from it have
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   proteins can be used in compositions for treating or preventing infection
CC   due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   bacteria. Computers, computer memory, computer storage medium or computer
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   further opportunities to find antigenic or immunogenic proteins which are
CC   more effective in vaccines than the outer membrane proteins currently
CC   used.
XX
SQ   Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

alignment_scores:
      Quality: 404.00          Length: 307
      Ratio: 2.138             Gaps: 8
Percent Similarity: 61.564    Percent Identity: 35.179

alignment_block:
US-09-357-675C-21 x AAF21607 ..

Align seg 1/1 to: AAF21607 from: 1 to: 349980

16 LeuCysProGlyLeuArgIleProGlnLeuSerValIleucySAIagInPr 32
||||| ||:::|| |::||::|| |
153960 TTGTGTAA.....CGGTTTCGCAAGTACCGTCATGGTAGCTATCG 154003
32 oAryg.....ProArqAlaMetAlaIleSerSerSerSercyscLueuP 47
||| ||::||::|| |:::|||| |
154004 GCGGAATATGCCCGCAAGTGCGGCAGGAAAAAGCAAGAATAATGACAAA 154053
47 roLeuValAlaValAlaCysGlnValTherThrProAspLysGlnGlnAsn 63
||||| ||::||::|| |:::|||| |
154054 TCAGAGTTCGCGCGCGTCACAGATGGTGTGCGGCGCGTCCGCGGAACAAC 154103
64 PheIysThrcysAlaGluLeuValArgGluAlaAlaIargIleuGylAcy 80
::: ::||| ||::||::|| |:::|||| |
154104 GTGCGCGCATGAAACGCCGTGTCGCGCAGGCGCGGAGAGAGGTGCGGA 154153
80 sIeuAlaPheLeuProGluAlaPheAspHeiIeAlaIargAproAlag 97
::: ||::||| ||::||::|| |:::|||| |
154154 TTGGGTGCTGCTCCCGAATATTGGGTGCTGATGGGCGCAACATATCCG 154203
97 IurThreunIsteuSerGluProLeu...GlyGIYIsIeuIenGluGln 112

```



```

154204 ACAAACTCGCGCTTGGCGCGGACGCGCTTCAACAGCGCA 154253
      :: ||| |||::||| ||| |||::|||
113 TTTThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 128
      ::|||::|||::|||::|||::|||::|||
154254 TTGAGCGAAGACGGCGAAGATGGCGGTGGTGGT...TTGCGCGGCGAC 154300
      ::|||::|||::|||::|||::|||
129 .....PheHisGluArgGlyGlnAspTrpGluGlnThrGlnL 141
      ||| |||
154301 TGTCCCGCTGCAAGCTCGCGAGCGGCT.....A 154329
141 ysiIeTyAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
      |||::|||::|||::|||::|||::|||
154330 AAGTGAATGATACCGCTGTGGTGTACGACGCGAGCGGTAGAGCGGG 154379
158 ThrTyArgLysThrHisLeuCysAspValGluLeuProGlyGlnGlyPr 174
      |||::|||::|||::|||::|||::|||
154380 CTGTACACCAAAATGCACCTC.....TTGCGTTTTCGGTTGGCGG 154423
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
      ::|||::|||::|||::|||::|||
154424 ACGCTATGCCAGACCCATACCATCCGCGCGGCGGATGCCGCACT 154473
191 alSerThrProAlaGlyLysIleGlyLeuAlaValCysTyAspMetArg 207
      ::|||::|||::|||::|||::|||::|||
154474 TGTGGCGAAGAGCGGTGGCGGCGGCGCATTTGTACAGATGCCG 154523
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluLeuThr 224
      |||::|||::|||::|||::|||::|||
154524 TTTTCCCAA.....TTTTCGACGCGCATTTGCCGTTTGCATTCAT 154567
224 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 241
      |||::|||::|||::|||::|||::|||
154568 GCTGCCCGCTGCGTTTACGACACGACGCGGCAAGCGCATTTGGAGTGC 154617
241 euleAlaArgAlaArgAlaIleGluThrGlnCysTyArgValAlaAla 257
      |||::|||::|||::|||::|||::|||
154618 TGTGCGCGCGCGGTGGCGTGAACCAATTTAGTGTGGCGGCGCA 154667
258 GlnCysGlyArgHisHisGlnLysArgAlaSerTyArgLysHisSerMet 274
      |||::|||::|||::|||::|||::|||
154668 CAGGCGCGTTTGCACGAAACGCGCGCGCATTTCCGACACAGCATGAT 154717
274 lValAspProTrpGlyThrValAlaAlaArgCysSerGluGlyProGly 291
      |||::|||::|||::|||::|||::|||
154718 TGTGATTCGCGGCGACGAGTGTGACGATTTGCCGAGGCGAAGCG 154767
291 euCysLeuAlaArgIleAspLeuAsnTyLeuArgGlnLeuArgArgHis 307
      |||::|||::|||::|||::|||::|||
154768 TTGTACGCGCAGACATGATGCCAACCGCTTGAACAGCGTCCGCAACGC 154817
308 LeuProValPheGlnHisArg 314
      |||::|||::|||::|||::|||::|||
154818 CTGCCCGCTTGAATACCGG 154838
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-embL/NA2000.DAT:AAA81490
seq_documentation_block:
ID AAA81490 standard; DNA; 1437668 BP.
XX
AC AAA81490;
XX
DT 04-DEC-2000 (first entry)
XX
XX N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; Menb; ds.
XX
OS Neisseria meningitidis.
XX
PN W0200022430-A2.

```

```

XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99NC-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Maignani V, Galeotti C, Mora M, Rattl G, Scarselli M, Scariato V,
PI Rappoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 866-1272; 1760pp; English.
XX

```

```

XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins. AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis Menb polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have also been tried but none have successfully
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

```

```

alignment_scores:
Quality: 404.00 Length: 307
Ratio: 2.138 Gaps: 8
Percent Similarity: 61.564 Percent Identity: 35.179

```

alignment\_block:

US-09-357-675C-21 x AAA81490 ..

Align seg 1/1 to: AAA81490 from: 1 to: 1437668

```

16 LeuGlyProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
|||::|||::|||::|||::|||::|||
453960 TTGTGTAAA.....CGGTTTTCGCAAAAGTACCGTCATGATGCTATCG 454003
32 oArg.....ProArgAlaMetAlaIleSerSerSerSerCysGluLeuP 47
|||::|||::|||::|||::|||::|||
454004 GCGGAATATTCGCGCAAGGTGCGCGCAGAAAGAGAGAAATATGACAAA 454053
47 rLeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnSn 63
|||::|||::|||::|||::|||::|||
454054 TCAGAGTTGCCCGCGTGCAGATGTGTGCGGCGTGTGCCGGAACCAAC 454103

```

```

64 PhelysThrcysalaGluLeuValArgGluAlaAlaArgLeuGlyAlaCys 80
   ::::::::::::::::::::
454104 GTGCGCCGCTGAAAGCGCTGTGCGACAGCGGCGGACGAGTGGCGGA 454153
   ::::::::::::::::::::
80 sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 97
   ::::::::::::::::::::
454154 TTGGGTCTCTGCTGCCAATATGTCGTGTCGTGCGGCAACATGACCG 454203
   ::::::::::::::::::::
97 lUrThrhLhLsLeuSerGluProLeu...GlyGlyLysLeuLeuGluGlu 112
   ::::::::::::::::::::
454204 ACAACACTGCGGCTTGGCGACCTTTGGCGGCGGACCGCTTACAGCGCA 454253
   ::::::::::::::::::::
113 TyrThrhGlnLeuAlaArgGlyCysGlyLeuTyrPheLeuSerLeuGlyGly 128
   ::::::::::::::::::::
454254 TTGAGCGAAACGCGAAAGATGCGCGCTGTGCTG...TTGCGCGGAGC 454300
   ::::::::::::::::::::
129 .....PheHsGluArgGlyGlnAspTyrPheGluGlnThrGln 141
   ::::::::::::::::::::
454301 TGTGCGCGCTGCAAGCTGCGAGCGGCT.....A 454329
   ::::::::::::::::::::
141 ySileTyrAsnCysHsIValLeuLeuAsnSerLysGlyAlaValAla 157
   ::::::::::::::::::::
454330 AAGTGAATGAAATACCGCTGTGTCGACGAGCGGCGGCTTACAGCGGCG 454379
   ::::::::::::::::::::
158 ThrTyrArgLysThrhLsLeuCysAspValGluIleProGlyGlnGlyPr 174
   ::::::::::::::::::::
454380 CTGTACCAACAAATGCACATC.....TTGCGTTTTCGCTTGGGCGGA 454423
   ::::::::::::::::::::
174 oMeCysGluSerAsnSerThrhMetProGlyProSerLeuGluSerPro 191
   ::::::::::::::::::::
454424 ACGTATGCGCAAGCGATCCATCCATCCGCGGCGGCGGAGTGGCGGACT 454473
   ::::::::::::::::::::
191 aLseThrhProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArg 207
   ::::::::::::::::::::
454474 TGTGCGAAGAGCGCTGCGGTCGCGGCGGCGGCTTGTTCAGATGTCGCG 454523
   ::::::::::::::::::::
208 PheProGluLeuSerLeuAlaAlaGlnAlaGlyAlaGluIleLeuTh 224
   ::::::::::::::::::::
454524 TTTCGCCGAA.....TTTTCGACGCCAGTTCGCGCTTGTTCAGATGAT 454567
   ::::::::::::::::::::
224 rTyProSerAlaPheGlySerIleThrhGlyProAlaHisThrIleVal 241
   ::::::::::::::::::::
454568 GCTGCCGCGCTGTTACGACACACGAGCGGCGGCGGCGGAGTGGGCTGC 454617
   ::::::::::::::::::::
241 euleuAlaAlaArgAlaIleGluThrhGlnCysTyrValAlaAlaAla 257
   ::::::::::::::::::::
454618 TGTGCGCGGCGGCTGCGCTGCAAAACCAATGTTCGTGTCGCGGCGCA 454667
   ::::::::::::::::::::
258 GlnCysGlyArgHisHsGluLysArgAlaSerTyrGlyHisSerMetVa 274
   ::::::::::::::::::::
454668 CAGGCGCGTTCGACGAAAGCGAGCGGCGGCGGCTTCGACACACATGAT 454717
   ::::::::::::::::::::
274 lValAspTyrProGlyThrhValAlaAlaArgCysSerGluGlyProGlyL 291
   ::::::::::::::::::::
454718 TGTGCAATCCGTGGCGGCGGCTGTTGACGATTCGCCGAGGCGCAAGCG 454767
   ::::::::::::::::::::
291 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 307
   ::::::::::::::::::::
454768 TTGTTCAGCGACAGATGATGCCAACCGCGCTGAACACGCTCCGCAACCGC 454817
   ::::::::::::::::::::
308 LeuProValPheGlnHisArg 314
   ::::::::::::::::::::
454818 CTGCGCGCGCTTGAATATCCGG 454838
   ::::::::::::::::::::

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AACT71716

seq\_documentation\_block:  
ID AACT71716 standard; cDNA; 1725 BP.

XX AAC71716;  
XX  
XX  
DT 08-FEB-2001 (first entry)  
XX

```

DE Human ORFX ORE2731 polynucleotide sequence SEQ ID NO:5461.
XX
KW Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipruritic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antihypertensive;
KW antiviral; antibacterial; antifungal; antipneumatic; antithyroid;
KW neurodegenerative disorder; cancer; proliferative disorder; hypertension;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
PI MPI: 2000-602362/57.
XX
DR P-PSDB; AAB42967.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 4641-4643; 5507pp; English.
XX
XX AACT74446 to AACT7606 encode the proteins given in AAB40237 to AAB4397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antipruritic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflammatory; antibacterial; antiviral; antifungal; antipneumatic;
XX antithyroid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1725 BP; 466 A; 368 C; 372 G; 517 T; 2 other:

```

alignment\_scores:  
Quality: 392.00 Length: 481  
Ratio: 1.858 Gaps: 14  
Percent Similarity: 43.867 Percent Identity: 24.324







**THIS PAGE BLANK (USPTO)**







LOCUS	AL520768	890 bp	MRNA	linear	EST 13-FEB-2001
DEFINITION	AL520768 L1L_NFL004_NBC2 Homo sapiens CDNA clone CS0DB02YD05 5				
ACCESSION	AL520768				
VERSION	AL520768.1				
KEYWORDS	GI:12784261				
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.				
JOURNAL	Full-length cDNA libraries and normalization				
COMMENT	Unpublished (2001)				
FEATURES	Genoscope - Centre National de Sequencage				
Source	BP 191 91006 EVRY cedex - France				
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
	Location/Qualifiers				
	1..890				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0DB02YD05"				
	/clone_lib="L1L_NFL004_NBC2"				
	/sex="male"				
	/tissue_type="neuroblastoma cells"				
	/lab_host="DH10B"				
	/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : filiang@lifetech.com URL : http://fulllength.invitrogen.com"				
BASE COUNT	191 a	251 c	241 g	206 t	1 others
ORIGIN					
alignment_scores:					
Quality:	1458.00	Length:		277	
Ratio:	5.302	Gaps:		0	
Percent Similarity:	99.278	Percent Identity:		99.278	
alignment block:					
US-09-357-675C-21 x AL520768	..				
Align seg 1/1	to: AL520768 from: 1 to: 890				
1 MetLeuGlyPheIleThrArgProHisAspGlyPheLeuSerLeuLeu	17				
60 ATGCTGGGGCATCATCCAGCGCCCTCTCTCAACATTCCTCTCTTG	109				
17 sProGlyLeuArgIleProGlnLeuSerValLeuGlySalagInProArg	34				
110 TCCTGGAGCTCGGATACCTCACTCTCACTACTTTGTGCTGACCGCAG	159				
34 roArAlaMetAlaIleSerSerSerSerGlyLeuProLeuValAla	50				
150 CCAGAGCCATGGCTATCTCTCTCTCTCTGGAACCTGCCCTGGGCT	209				
51 ValCysGlnValThrSerThrProAspLysGlnGlnAspPheLysThr	67				
210 GTGTGTCACAGTACATCGACGCCAGACAAACAGAACTTTAAACATG	259				
67 salagLeuValArgGlnAlaAlaArgLeuGlyAlaCysLeuAlaPhe	84				
260 TGCTGAGTGTGTCGAGAGCTGCAGACTGGGTGCTCTGCGCTTCC	309				
84 eurProGlnAlaPheAspPheIleAlaArgAspProAlaIuThrLeuHis	100				

310 TGCCTGAGGCATTTCGACTTCATTTGCACGGGAGCCCTCAGAGACGCTACAC 359

101 LeuSerGluProLeuGlyGlyLeuLeuGluGluTyrThrGlnLeuAl 117  
|||||  
360 CTGTGTGAACCACTGGGTGGGAAACTTTTGGAAGATATACACCAGCTTGC 409

117 aargGluCysGlyLeuTTrPLeuSerLeuGlyGlyPheHisGluArgGlyG 134  
|||||  
410 CAGGGAATGTGGACTGTGGCTGCTCCCTTGGGTGGTTTCCATGAGCCGTGGCC 459

134 LAspPrpGluGlnThrGlnLysLLeTyrAsnCyHisValLeuLeuAsn 150  
|||||  
460 AAGACTGGGAGCACTCAGAAAATTTACAATTTGTACGTGCTGTGAAC 509

151 SerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAspVa 167  
|||||  
510 AGCAAAAGGGGCACTAGTGGCCACTTACAGAGACACATCTGTGTACGT 559

167 LgIuIleProGlyGlyGlyProMetCysGluSerAsnSerThrMetProG 184  
|||||  
560 AGAGATTTCCAGGGCAGGGGCTATGTGTAAAGCAACTTACCATCTCTG 609

184 LProSerLeuGluGluSerProValSerThrProAlaGlyLysLLeGlyL 200  
|||||  
610 GGCCGAGCTTGTAGTACCTGTGACGACACACAGCAAGATGGTGTCTA 659

201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217  
660 GCGGTGTGTATGACATGGCGTTCCCTGAACCTCTCTGGCANTGGCTCA 709

217 naLaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerLleThrG 234  
|||||  
710 AGCTGAGAGAGAGATCTTACCTATCTTCAGCTTTGGATCCATTAACAG 759

234 LProAlaHisTrpGluValLeuLeuArgAlaArgAlaLleGluTrpGln 250  
|||||  
760 GCCCACCCCACTGGGAGGTGTGCTCGGGCCCTGTCTATCGAAACCCAG 809

251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluIuAsArgAl 267  
|||||  
810 TGTGTATGTATGGCAGCAGCAGCAGTGTGGACGCCACATGAGAGAGAGC 859

267 aSerTyrGlyHisSerMetValValAspPro 277  
|||||  
860 AAGTTATGGCCACAGCATGTGTGGKAGACCCC 890

seq\_name: gb\_htc:AK002265

seq\_documentation\_block:

LOCUS AK002265 1374 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610007B07:nitrilase 1, full insert sequence.

ACCESSION AK002269

VERSION AK002269.1 GI:12832130

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:0610007B07.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2 (sites)

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL	PUBMED 204929374
REFERENCE	110421929
AUTHORS	3 (sites) Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishikawa,K., Katsunuma,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasaiwagi,K., Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yokeda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawal,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL	PUBMED 20530913
MEDLINE	11076681
PUBMED	4 (sites) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
REFERENCE	Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
TITLE	5 (bases 1 to 1374)
JOURNAL	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Butt,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hangaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiroka,T., Horii,F., Hume,D., Imocani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishikawa,K., Nomura,K., Numataki,R., Ono,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Satoh,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
JOURNAL	Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGACGGCCGCCACACTGCATCTTTTCTTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Ret = 5.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAAGAAGAGAGATCCAGAGCTCATTAATTAAATTAACCACCCCCCCC 3']. cDNA was cleaved with XhoI and SfiI. Cloning sites, 5' end: SfiI; 3' end: XhoI. Host: SOLR.
FEATURES	Location/Qualifiers 1..1374 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="MGD:1891931" /db_xref="taxon:10090" /clone="0610007B07" /sex="male" /tissue_type="kidney" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult"
SOURCE	1..1374 /misc_feature 1..1374 /gene="Ntl1"

/note="data source:MGD, source key:MG1.1350916  
evidence:ISS  
nitrilase 1"

BASE COUNT	330 a	339 c	342 g	343 t
ORIGIN				

alignment\_scores:  
 Quality: 1458.00      Length: 327  
 Ratio: 4.780      Gaps: 3  
 Percent Similarity: 93.272      Percent Identity: 83.792

alignment block:  
 US-09-357-675C-21 x AK002269 ..

Align seg 1/1 to: AK002269 from: 1 to: 1374

```

2 LauglyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
|||||.....:|||||
119 CTGGGCTTCATCACCAGCGCTCTCCACCA.....CTCTGTGTAC 159

18 oclyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgPro 35
|||.....:|||||
160 CGGATACCGATTACTTCGAACCCACAGTACTTGTACTCAGCCAGGCCA 209

35 rGAlaMeTalIleSerSerSerSerCysGluLeuProLeuValAlaVal 51
|||||.....:|||||
210 GACACATGTCC...TCATCACTTCTCTGGAGAGCTGCCCTGGGGCTGTG 256

52 CysGlnValThrSerThrProAspProAspGlnAsnPhySerCysAl 68
|||||.....:|||||
257 TGCCAGGTACATCAACACCAACCAAGCAAGACATTTAAACATGTGC 306

68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeu 85
|||||.....:|||||
307 TGA GTTGGTTCAAGAGCGTCCAGACTGGGTGCTTCCTGGCCTTTCTGC 356

85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
|||||.....:|||||
357 CTGAGGCAATTGACTTATTTATGACAGAAACCTGCCAGACATTAACCTC 406

102 SerGluProLeuGlyGlyLeuLeuGluGluGlyThrGlnLeuAlaArg 118
|||||.....:|||||
407 TCCGAACCCCTGAATGGGAGATCTTTGGGCCAATATAGCCAGCTCCAG 456

118 gGluCys_GlyLeuThrPheLeuSerLeuGlyGlyPheHisGluArgGlyGln 134
|||||.....:|||||
457 GGAATCTTGATTCAGTTCGTCTTGGGGGCTTCCACAGCGTGGCCAA 506

135 AspTPGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSe 151
|||||.....:|||||
507 GACTGGGACAGAAATCAGAAATCTCAATTGTCATGATGCTTTTGAACAG 556

151 rlySGlyAlaValAlaAlaThrTyrArgGlyThrHisLeuCysAspValG 168
|||||.....:|||||
557 CAAGGATTCAGTAGGGCCAGTTACGGAAGACACATTCGTGTGAGATGAG 606

168 lAlaIleProGlyGlnGlyPrometCysGlnSerAsnSerThrMetProGly 184
|||||.....:|||||
607 ACATCCAGGTACAGGGGCCGATGAGAGAAAGCAACATATACCAAGCTGGA 656

185 ProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAl 201
|||||.....:|||||
657 GGCACCTCTTGACCCACTGTCAAGACACCGGCTGGCAAGTTGGTTACAC 706

201 aValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnA 218
|||||.....:|||||
707 AATCTGTATGTACATCGCGGTCCCTGAACTTCTTTGAAATTGGCTCAAG 756

218 lAglyAlaGluLeuLeuThrTyrProSerValaPheGlySerIleThrGly 234
|||||.....:|||||
757 CTGGGGCAGAAATACTTACTTATCTTCAGCCCTTATGATCTGTATACAGT 806

```

235 ProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCys 251  
|||||  
807 CCGAGCCACATGGAGTGTGCTGGGGGCGCCGCTTGAATCTCGTGTG 856  
|||||  
251 sTrValValAlaAlaAlaGlnCysGlyArgHisHisGlnLysArgAlaAs 268  
|||||  
857 CTATGTAAATAGCAGCAGCCAGTGTGGACCCACCATGAACAAGACAA 906  
|||||  
268 eTrArgGlyHisSerMetValValAspProTrpGlyThrValAlaAlaArg 284  
|||||  
907 GTTATGGCCATAGCATGTGGTTGACCCGTGGGCGACAGTGTGGCCGC 956  
|||||  
285 CysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLysAsnTyrLe 301  
|||||  
957 TGTCCGAGGAGACAGGCTCTGCTGCTGCGAATTCATCTCCACTTCT 1006  
|||||  
301 uArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspL 318  
|||||  
1007 ACAACAGATCGCCACACCTGCTGTTCACAGACAGACAGCTGACC 1056  
|||||  
318 eUTrGlyAsnLeuGlyHisProLeuSer 327  
|||||  
1057 TCTATGGCAGTCTGGGTCAATCTCTCT 1085  
|||||  
seq\_name: gb\_est1.AL529152

seq\_documentation\_block:  
LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL529152 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CS0DD004YM06 5  
prime, mRNA sequence.  
ACCESSION AL529152  
VERSION AL529152.1 GI:12792645  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 844)  
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
Source  
1..844  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DD004YM06"  
/clone\_id="LTI\_NFL001\_NBC4"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed  
by Life Technologies, a division of Invitrogen 9800 Medical Center  
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
8371 Email : filangel@life.techn.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 182 a 238 c 235 g 189 t  
ORIGIN  
alignment\_scores:  
Quality: 1413.00 Length: 263  
Ratio: 5.373 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-357-675C-21 x AL529152 ..  
Align seq 1/1 to: AL529152 from: 1 to: 844

65 LysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLe 81  
|||||  
1 AAACAGTGTCTGTGAGTGTGTTGAGAGGCTGCCAGACTGGGTGCTGCT 50  
|||||  
81 uAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGlu 98  
|||||  
51 GCGTTTCTCCCTGAGGCGAATTTCATTGATTCACAGGAGACCTCCAGAGA 100  
|||||  
98 hTrLeuHisLeuSerGluProLeuGlyLysLysLeuGluGluTyrThr 114  
|||||  
101 CGCTACACTGTCTGAACCACTGGGTGGAAACTTTTGGAAATTAACACC 150  
|||||  
115 GlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisG 131  
|||||  
151 CAGCTTGCCAGGGAATGTGACTGTGCTGCTGTGGGTGTTCCATGA 200  
|||||  
131 uArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisVal 148  
|||||  
201 GCGTGGCCAAAGACTGGAGCAGACTCAGAAATCTAACAATGTGCACGTGC 250  
|||||  
148 eUeUAsnSerLysGlyAlaValAlaValAlaThrTyrArgLysThrHisLeu 164  
|||||  
251 TGTGAACAGCAAAAGGGCAGTACTGGCCACTTACAGAAACACACTGTG 300  
|||||  
165 CysAspValGluLeuProGlyGlnGlyProMetCysGluSerAsnSerTh 181  
|||||  
301 TGTACCTAGAGATTCCAGGCGAGGGCCCTATGTGTGAACCACTGTAC 350  
|||||  
181 rMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 198  
|||||  
351 CATGCTGGGCGCCAGTCTGTGAGTCACTGCACGACAGACACAGCAGCAAGA 400  
|||||  
198 legGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAla 214  
|||||  
401 TTGGTCTAGCGTCTGCTATACATGCGGGTCCCTGAACCTCTCTGGCA 450  
|||||  
215 LeuAlaGlnAlaGluAlaGluIleLeuThrTyrProSerAlaPheGlySe 231  
|||||  
451 TTGGCTCAAGCTGAGCAGACAGATCTTACCTATCTTCAGCTTTTGATTC 500  
|||||  
231 rIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleG 248  
|||||  
501 CATTACAGGCCACAGCCCACTGGAGGTGTGCTGCGGGCCCGTCTATACG 550  
|||||  
248 lUTrpGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGlu 264  
|||||  
551 AAACCCAGTGTATGTAGTGGCAGCAGCAGATGTGAGCGCCACCATGAG 600  
|||||  
265 LysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVa 281  
|||||  
601 AAGAGACCAAGTTATGGCCACAGCATGTGTATACCCCTTGGGAACAGT 650  
|||||  
281 lValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgTLeAspL 298  
|||||  
651 GGTGGCCCGCGTCTGTGAGGGGCCAGGCGCTGCTGCCCGGAAATGAGCC 700  
|||||  
298 eUAsnTyrLeuAlaGlnLeuAlaArgArgHisLeuProValPheGlnHisArg 314  
|||||  
701 TCAACTATCTGCGACAGTGTGGCGGACACCTGCTGTGTTCCAGACCCGC 750  
|||||  
315 ArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 327  
|||||  
751 AGGCTGACCTCTATGGCAATCTGGGTACCCACTGTCT 789  
|||||  
seq\_name: gb\_est2:BI822844  
seq\_documentation\_block:  
LOCUS BI822844 928 bp mRNA linear EST 04-OCT-2001

DEFINITION 603040166F1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5180897 5', mRNA sequence.

ACCESSION BI822844

VERSION BI822844.1 GI:15934394

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

AUTHORS NIH-MGC http://mhc.nci.nih.gov/.

TITLE 1 (bases 1 to 928)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: L1AM11451 row: d column: 18

High quality sequence stop: 822.

Location/Qualifiers

1..928

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5180897"

/clone\_1lb="NIH\_MGC\_115"

/lab\_host="DH10B"

/note="Organ: pooled brain, lung, testis; Vector: PCMV-SORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."

BASE COUNT 198 a 256 c 259 g 214 t 1 others

ORIGIN

alignment\_scores:

Quality: 1357.50 Length: 295

Ratio: 4.848 Gaps: 4

Percent Similarity: 94.915 Percent Identity: 93.559

alignment\_block:

US-09-357-675c-21 x BI822844 ..

Align seg 1/1 to: BI822844 from: 1 to: 928

1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeu 17

|||||

39 ATGCTGGGCTTCATCACAGGCTCCGACAGATTCCTGTCCTTGTG 88

|||||

17 sProGlyLeuArgIleProGlnLeuSerValLeuGlyAlaGlnProArg 34

|||||

89 TCGTGAAGCTCCGGATACCTCACTCAGACTTGTGCTCACCACGCG 138

|||||

34 roArGAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuValAla 50

|||||

133 CCGAGAGCGAAGGCTATCTCTCTCTCTCCGCACTGCCCTCGTGCGCT 188

|||||

51 ValCysGlnValIleThrSerThrProAspLysGlnGlnAsnPheLysThr 67

|||||

189 GGTGTCCAGGTACATGACGCGACAGACAGACAGAACTTTAAACATG 238

|||||

67 sAlaGlnLeuValIleArgGlnAlaIleArgLeuGlyAlaCysLeuAlaPhe 84

|||||

239 TCGTGAAGCTGGTTCGAGAGGCTGCCAGACTGGGTGGTCCGCTGGCTTTC 288

84 eupProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThrLeuHis 100

|||||

289 TCGCTGAGGCATTGTGACTTCATTGSCAGGAGCCCTGCAGAGAGCTACAC 338

|||||

101 LeuSerGlnProLeuGlyGlyLysLeuLeuGlnGlnLysThrGlnLeuAl 117

|||||

339 CTGCTGAAACCACTGCGGTGGGAACTTTGGAAGAAATACACCACGAGCTGC 388

|||||

117 aArgGlnCysGlyLeuThrProLeuSerLeuGlyGlyPheHisGlnArgLys 134

|||||

389 CAGGCAATGTGGACTGTGGCTGTGGTGGTGGTGTTCATGAGCGGTGGC 438

|||||

134 InAspTrpGlnGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150

|||||

439 AAGAGCTGGAGCAGACTCACAANAATACAAATGTCTACGCTGCTGTAAC 488

|||||

151 SerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeuGlyAspVal 167

|||||

489 AGCAAAAGGGCAGTACGTGGCCACTTACAGAGAACACATCTGTGTGAGT 538

|||||

167 LglnIleProGlyGlnGlyProMetCysGlnSerAsnSerThrMetProG 184

|||||

539 AGAGATTCCAGGCAG. GGGCTATGTGTGAAGCACTTACCATGTCGT 587

|||||

184 lYpProSerLeuGlnSerProValSerThrProAlaGlyLysIleGlyLeu 200

|||||

588 GGCCCACTCTTGAGTCACTGTGACGCCCAACAGCAGCAAGATTTGGCTTA 637

|||||

201 AlaValCysTyrAspMetArgPheProGlnLeuSerLeuAlaLeuAlaG 217

|||||

638 GCTGTGCTATGATGACATGCGGTTCCCGAAGCTCTCTGCAATGGCTCA 687

|||||

217 naIaGlyAlaGlnIleLeuThr TyrProSerIlePheGlySer... Ile 232

|||||

688 AGCTGAGGACAGATTAATCTTATCTTCACTTGTGGATCCCATATA 737

|||||

233 ThrGlyProAlaHisTrpGlnValLeuLeuArgAlaArgAlaIleGlnThr 249

|||||

738 CAGGGCCCAACCCACATGGAGAGTGTGCTCGGGGNCGGCTATTCGAAC 787

|||||

249 rGlnCysTyrVal ValAlaAlaAlaGlnCysGly ArgHisHisGlnLys 265

|||||

788 CCAATGCTATGTATGATGGCCGACACAGATGTGAACGCCACCATGAGAA 837

|||||

265 sArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVal 282

|||||

838 GAGAGC. AGTTATGGGCACAGCATGGGTGAACCCCTGGGGACAGGGGG 886

|||||

282 alAlaArgCysSerGlnGlyProGlyLeu 291

|||||

887 CCCGTGCTCGAGGCGCAGCGCTGGCTTG 915

|||||

seq\_name: gp\_est2:BI769604

seq\_documentation\_block:

LOCUS BI769604 846 bp mRNA linear EST 25-SEP-2001

DEFINITION 603054985F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5204469 5', mRNA sequence.

ACCESSION BI769604

VERSION BI769604.1 GI:15761182

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

AUTHORS NIH-MGC http://mhc.nci.nih.gov/.

TITLE 1 (bases 1 to 846)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: L1M11512 row: 3 column: 22  
 High quality sequence stop: 838.  
 Location/Qualifiers

## FEATURES

source

1..846  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5204469"  
 /lab\_host="NIH\_MGC\_122"  
 /lab\_host="DH10B"  
 /note="Organ: pooled lung and spleen; Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH MGC Library."  
 BASE COUNT 179 a 240 c 226 g 201 t  
 ORIGIN

## alignment\_scores:

Quality: 1308.00 Length: 264  
 Ratio: 4.992 Gaps: 2  
 Percent Similarity: 99.242 Percent Identity: 99.242

## alignment\_block:

US-09-357-675c-21 x B1769604 ..

Align seg 1/1 to: B1769604 from: 1 to: 846

1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeu 17  
 |||||||  
 59 ATGCTGGGCTTCATCAACAGGCTCCACAGATTCCTGCTTCGTG 108  
 |||||||  
 17 SPGCGLYLeuArgIleProGlnLeuSerValLeuGlyAlaGlnProArg 34  
 |||||||  
 109 TCCTGGACTCCGATACCTCACTCAGACTTGTGCTCAGCCAGCG 158  
 |||||||  
 34 roArGAlaMetAlaIleSerSerSerSerSerSerSerSerSerSer 50  
 |||||||  
 159 CCAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 208  
 |||||||  
 51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThr 67  
 |||||||  
 209 GTGTGCAGTACATCGACGACGACGACGACGACGACGACGACGAC 258  
 |||||||  
 67 salAgluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84  
 |||||||  
 259 TGCTGAGCTGGTTCGAGAGGCTCCAGACTGGGTGCTGCTGCTTTC 308  
 |||||||  
 84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
 |||||||  
 309 TGCTTGAAGCATTTGACTTATTCACAGGACCTGACAGAGCTACAC 358  
 |||||||  
 101 LeuSerGluProLeuGlyLysLeuLeuGluGluThrThrGlnLeuAla 117  
 |||||||  
 359 CTGTCTGAACCACTGGGTGGGAAACTTTGGAGAAATACACCACTTGC 408  
 |||||||  
 117 aaArgLysGlyLeuThrLeuSerLeuGlyGlyPheHisGluArgGly 134  
 |||||||  
 409 CAGGAGATGTGGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 458  
 |||||||  
 134 InaSPTrpGluGlnThrGlnLysIleTyrAsnGlyHisValLeuLeuAsn 150

|||||  
 459 AAGACTGGAGCAGACTCAGAAAATCTACAAATGTCTACGCTGCTTAAC 508  
 |||||||  
 151 SerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeuCysAsp 167  
 |||||||  
 509 AGCAAGAGGCGACTAGTGGCCACTTACAGAGACACATCTGTGTGACGT 558  
 |||||||  
 167 IgLulleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPro 184  
 |||||||  
 559 AGAATTCCAGGGCAGGGGCTTATGTGTGAAGCACTTACATGCTGCTG 608  
 |||||||  
 184 LysProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
 |||||||  
 609 GGGCCATCTTGAAGTACCTGCTGACGACACACAGCAGGCAAGTTGCTCA 658  
 |||||||  
 201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 217  
 |||||||  
 659 GCTGTCACTGATGACATGCGGTTCCTCGAAGCTCTCTGCGCATTTGCTC 708  
 |||||||  
 217 InaIaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233  
 |||||||  
 709 AAGCTGGAGGAGAGATCACTTATCTTCTGAGCTTTGGATCCATTACA 758  
 |||||||  
 234 GlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrG 250  
 |||||||  
 759 GGGCCAGC.CACTGGAGGTGTTCCTGCGGGC.GTGTCTATCGAAGACCA 806  
 |||||||  
 250 nCysTyrValAlaAlaAlaAlaGlnCysGlyArgHis 262  
 |||||||  
 807 GTGCTATGTAGTGGCAGCCAGCACAGTGTGACGCCAC 844  
 |||||||

seq\_name: gb\_est2:BG436916

seq\_documentation\_block:

LOCUS BG436916 793 bp mRNA linear EST 14-MAR-2001  
 DEFINITION 602488424F1 NIH\_MGC\_18 Homo sapiens CDNA clone IMAGE:4620892 5',  
 mRNA sequence.  
 ACCESSION BG436916  
 VERSION BG436916.1 GI:13343422  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 793)  
 NIH-MGC <http://mhc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: DCTD/DTF/Gazdar  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: L1CMI382 row: 0 column: 05  
 High quality sequence stop: 767.  
 Location/Qualifiers

## FEATURES

source

1..793  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4620892"  
 /clone\_lib="NIH\_MGC\_18"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pORF7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and



```

|||||
178 GTGTGCAGTAACTGACGACGACGAGAACAAVAGAACTTTAAACATG 227
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
228 TGTGTAGCTGTTCGAGAGGCTGCCACACGTGGGCTGCTGCTTCC 277
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||
278 TGCTGTAGGACTTTGACTTATTCATGGGACCCCTGCAGAGACCTTACAY 327
101 LeuSerGluPro..LeuGlyGlyLeuLeuGluGluTyrThrGlnLeuA 117
|||||
328 CTGTCTGAACCTACTGGGTGGGAAACTTTTGAAGATATACACGACTTK 377
117 laArgGlu..CysGlyLeuTyrPleuSerLeuGlyGlyPheHisGluArgL 133
|||||
378 CCAAGGATATGTGACTCTKCTGCTTCTGTGGGTGCTTCCATGAGCTGG 427
133 yGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuA 150
|||||
428 CYMGAGCTGGGACGACACTYAGAAATCTCAAAATGTCACTGCTGCTGA 477
150 snSerLysGlyAlaValAlaAlaThrTyrArgLysThrHisLeuCysAsp 166
|||||
478 ACAGCAAGGGCAGTAGTGGCCACTTACAGGAAGACATCTGTGTGAC 527
167 ValGluLeuProGlyGlnGlyPrometCysGluSerAsnSerThrMetPr 183
|||||
528 GTGAGACTTCACAGGCGGCGCTATGTGAAAAGCAACTCTACCAAGCC 577
183 oGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyL 200
|||||
578 TGGGCCCAAGCTTGTAGTCACTGTGCACACACAGCAGGCAAGATTGTC 627
200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 216
|||||
628 TACTGTCTGTATGACATGCGGTCCCTGCACCTCTCTCRCATTGGCT 677
217 GlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIle..T 233
|||||
678 CMAGCTGGGCCACAGAMACTTMMVCTTMCRCCTTTGGCWCACATTTA 727
233 hrcGlyProAlaHisTrp..GluValLeuLeuArgAlaArgAlaIleGluTh 249
|||||
728 CAGGCCCAAGCCCACTGGGCAAGTGTGCTCGGGCCCTGCTCWCMAAM 777
249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
|||||
778 CCAAGCTCTMMGTAGTGCWGCACACACAGTGTGAGCGCCACMATGAGAAGA 827
266 rGlnAspTyrGlyHisSerMetValValasp 276
|||||
828 GAGCAAGTTATGGCCACACGATGTGTAGAC 859
seq_name: gp_est12.BG762506

seq_documentation_block:
LOCUS BG762506 764 bp mRNA linear EST 15-MAY-2001
DEFINITION 602733954F1 NIH_MGC_49 Homo sapiens CDNA clone IMAGE:4859292 5',
RNA sequence.
ACCESSION BG762506
VERSION BG762506.1 GI:14073159
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 764)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC).
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

```

Email: cgaabs-f@mail.nih.gov
Tissue Procurement: ATCC/DCMD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1714 row: 1 column: 13
High quality sequence stop: 749.
FEATURES
source
1..764
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ made by skin; Vector: pORF7; Site.1: XhoI; Site.2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 162 a 218 c 201 g 183 t
ORIGIN
alignment_scores:
Quality: 1218.00 Length: 240
Ratio: 5.183 Gaps: 2
Percent Similarity: 97.917 Percent Identity: 97.917
alignment_block:
US-09-357-675C-21 x BG762506 ..
Align seg 1/1 to: BG762506 from: 1 to: 764
1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuGly 17
|||||
49 ATGCTGGGCTTTCATCCACGAGCCCTCCACAGATTCTGTCCCTTCTGTG 98
|||||
17 sProGlyLeuArgIleProGln..LeuSerValLeuCysAlaGlnProArg 33
|||||
99 TCCTGGACTCCGGATACCTCTAATCTCAAGACTTGTGGCTCAGCCAGG 148
|||||
34 ProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAl 50
|||||
149 CCCAGACCAATGGCTATCTCTCTCTCCCTCGCAACTGCCCCCGGGGCG 198
|||||
50 aValCysGlnValHisSerThrProAspLysGlnGlnAsnPheLysThrC 67
|||||
199 TGTGTGCCAGGTAACTATGATTCATTCACGAGGACCCCTGCAGAGAGCTT 248
|||||
67 ySalGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 83
|||||
249 GTGTGTAGCTGTTCGAGAGGCTGCCACAGACTGGGTGCTGCTGCTTTC 298
|||||
84 LeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuH 100
|||||
299 CTGCTGAGGCAATTTGACTTCATTCACGAGGACCCCTGCAGAGAGCTTACA 348
|||||
100 sLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuA 117
|||||
349 CTTGTCTGAACCTGGGTGGGAACTTTTGGGAGAAATATACACCAAGCTTG 398
|||||
117 laArgLysCysGlyLeuTyrPleuSerLeuGlyGlyPheHisGluArgGly 133
|||||
399 CCAAGGATATGTGACTCTGCTGCTTCTGTGGGTGCTTCCATGAGACGTGGC 448

```

```

134 glnasptprgluglnthrlnlyslletryasnrcysnhsvalleuleuas 150
|||||
449 CAAGACTGGGAGAGACTCAGAAATCTACAATTGCACGTGCTGTGAA 498
|||||
150 nserlysglyalavalalathrttyrarglysthrlisleucysaspy 167
|||||
499 CAGCAAGGGGGCGTAGTGGCCCTTACAGGAAGACACATCTGTGTGACG 548
|||||
167 alglulieproglynglynglyprometcysgluserasnserthmetpro 183
|||||
549 TAGAGATTCCACGGCAGGGGCTATGTGTGAAGCAACTCTACCATGCTC 598
|||||
184 glyproserleugluserprovalserthrproalaglylsileglyle 200
|||||
599 GGCCCAAGTCTTGACGTACCTGTGCACACACGACGAGCAAGTTGTGTC 648
|||||
200 ualavalcystryaspmetcargpheprogluleuserleualalaleu 217
|||||
649 ACCTGTCTCTATGACATCGGTCCCTGAACCTCTCTGGCATTTGGCTC 698
|||||
217 lualaglyalaglyluleuthrttyrproseralapheglyserilethr 233
|||||
699 AAGCTGACACAGAGATCTTACTATCTTCATCTTGGATTCATTACA 748
|||||
234 glypro.alaahstyp 238
|||||
749 GGCCCAAGGGCCACTGG 764
|||||

seq_name: gb_est2:BG703176

seq_documentation_block:
LOCUS BG703176 790 bp mRNA linear EST 07-MAY-2001
DEFINITION 602686082P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4818405 5',
mRNA sequence.
ACCESSION BG703176
VERSION BG703176.1 GI:13975252
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 790)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LHAM10720 row: d column: 22
High quality sequence stop: 766.
Location/Qualifiers
1..790
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4818405"
/clone_1lb="NIH_MGC_95"
*/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pBluescriptpr (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National

```

```

BASE COUNT 167 a 228 c 206 g 189 t
ORIGIN
alignment_scores:
Quality: 1206.00 Length: 238
Ratio: 5.132 Gaps: 2
Percent Similarity: 98.739 Percent Identity: 97.899
alignment_block:
US-09-357-675c-21 x BG703176
Align seg 1/1 to: BG703176 from: 1 to: 790

1 MetleuGlyPheIleThrArgProPheHisArgPheLeuSerleuency 17
|||||
79 ATGCTGGGCTTCATTCACACAGGCTCTCTCACAGATTCCTGCTGTGTG 128
|||||
17 sproglyleuargyleproglinleuservalleucysalaglnproargp 34
|||||
129 TCCTGGACTCCGGATACCTCACTCTCAGTACTTTGTGTCACAGCCAGG 178
|||||
34 roargalamealalieserSerSerSerCysGluLeuProleuValAla 50
|||||
179 CCAGAGCCATGGCTATCTCTCTCTCTCCGCGAACCTGGCCCTGTGGCT 228
|||||
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThr 67
|||||
229 GTGTGCGAGGTAAATCGACAGCCAGCAAGACACAGAACTTTAAACATG 278
|||||
67 salaglyleuvalarglyalalarglyleuglyalacysleualaphel 84
|||||
279 TGCTGAGCTGGTTGAGAGAGCTCCAGACTGGGTGCTGCTGCTTTC 328
|||||
84 euproglualaphasphellealalargaspProalagluthrleuHis 100
|||||
329 TGCTGAGGCAATTCATTCATTCAGCAGGAGCCCTGCAGAGAGCTAAC 378
|||||
101 leusergluleuProleuglylyslleuGlnGluGlyThrThrGlnleu 117
|||||
379 CGGTGGAACCACTGGGTGGGAACTTTGGAGAAATACACCCAGCTTGC 428
|||||
117 aargglyCysGlyLeuThrLeuSerleuGlyGlyPheHisGlyArgGly 134
|||||
429 CAGGAAATGTGACCTGGCTGTGCTTGGGTGTTCCATGAGCATGGCC 478
|||||
134 lnasptprgluglnthrlnlyslletryasnrcysnhsvalleuleuas 150
|||||
479 AAGACTGGGAGCAGACTCAGAAATCTACAATTGTGCATGCTGCGAAG 528
|||||
151 serlysglyalavalalathrttyrarglysthrlisleucysaspy 167
|||||
529 AGCAAGGGGCAAGTAGTGGCCACTTACAGGAAGACACATCTGTGTAGCG 578
|||||
167 lglulieproglynglynglyprometcysgluserasnserthmetpro 184
|||||
579 AGAGATTCCAGGCGAGGGCTATGTGTGAAGCAACTCTACCATGCTCG 628
|||||
184 lyproserleugluserprovalserthrproalaglylsileglyleu 200
|||||
629 GGCCCAAGTCTTGACGTACCTGTGCACACACGACGAGCAAGTTGTGTC 678
|||||
201 AlavalCysTyr. Aspmetcargpheprogluleuserleualalaleu 217
|||||
679 GCTGTCTGATTGACATCGGTCCCTGAACCTCTCTGGCATTTGGCTC 728
|||||
217 lualaglyalaglyluleuthrttyrproseralapheglyserile. 233
|||||
729 AAGCTGACACAGAGATCTTACTATCTTCATCTTGGATTCATTAC 778
|||||
233 rglproala 236
|||||

```



779 AGGCCAGACC 788  
seq\_name: gb\_est1:AL520767  
seq\_documentation\_block:  
LOCUS AL520767 992 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL520767 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CS0DB002YD05 3  
prime, mRNA sequence.  
ACCESSION AL520767  
VERSION AL520767.1 GI:12784260  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 992)  
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
source  
1. 992  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DB002YD05"  
/clone\_lib="LTI\_NFL004\_NBC2"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed  
by life technologies. Contact: Feng Liang life  
Technologies, a division of Invitrogen 9800 Medical Center  
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
8371 Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"  
BASE COUNT 229 a 267 c 268 g 222 t 6 others  
ORIGIN  
alignment\_scores:  
Quality: 1201.50 Length: 236  
Ratio: 5.179 Gaps: 3  
Percent Similarity: 98.305 Percent Identity: 96.610  
alignment\_block:  
US-09-357-675c-21 x AL520767/rev ..  
Align seg 1/1 to reverse of: AL520767 from: 1 to: 992  
94 ASPProLaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuLe 110  
|||||  
983 GACCTCTCAGAGACGCTACACCTCTGAACTGCTGGTGGAAC...TT 937  
|||||  
110 uGluGlu.TyrThrGlnLeuAlaArgGlyGlyGlyLeuTrpLeuSer 126  
: : : : :  
936 TTGAGATACACCCAGCTGCCAGGAATGAGCTGCTGCTGCTG 887  
|||||  
127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnGly 143  
|||||  
886 GGTGCTTCATGAGCGGTGGCAAGACTGGAGAGACTGAGAAATGTA 837  
|||||  
143 rAsnCysHisValLeuLeuAsnSerGlyAlaValAlaThrTyr 160  
|||||  
836 CAATTGTCACGCTGCTGTAAGCAAGGGGAGTAAGTGGCCACTTACA 787  
|||||  
160 rGlyStnHisLeuGlyAspValGluIleProGlyGlnGlyProMetCys 176

786 GGAAGACACATCTGTGTACGTAGAGATTCCAGGGGAGGGGCTATGTCT 737  
|||||  
177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193  
|||||  
736 GAAAGCAACTCTACCATCCCGGGCCGAGCTTGTGTCACCTGTCCAGCAC 687  
|||||  
193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetLrpPheProG 210  
|||||  
686 ACCAGCAGGCAAGATTGGCTAGCTGTCTGCTATGACATTGGGCTTCCTG 637  
|||||  
210 luleuSerLeuAlaLeuAlaGlnAlaGluIleLeuThrTyrPro 226  
|||||  
636 AACCTCTCTGCAATGTGGCTCAAGCTGGAGAGATATCTTACCATCT 587  
|||||  
227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243  
|||||  
586 TCAGCTTTTGATCCATATACAGGCCACACCTGGAGGTGTGCTGCG 537  
|||||  
243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaGlnCys 259  
|||||  
536 GGCCGTCGTATCGAAACCCAGTGTATGTAGTGCACGACACAGTGT 487  
|||||  
260 GlyArgHisGlyLysArgAlaSerTyrGlyHisSerMetValAlaVal 276  
|||||  
486 GGAGCCCAACCATGAGAGAGCAAGTATATGCCACACATGCTGTGAGA 437  
|||||  
276 pProTrpGlyThrValAlaAlaArgCysSerGluGlyProGlyLeuGly 293  
|||||  
436 CCCCTGGGAGACAGTGTGGCCGCTGCTGAGGGGCGGACCCCTGCTCC 387  
|||||  
293 euAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgHisLeuPro 309  
|||||  
386 TTGGCCGATGACCTCACTATGCGAGAGTGGCGCCACCTGCT 337  
|||||  
310 ValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGlyHisProLe 326  
|||||  
336 GTGTTCACGACACCGCAGGCTGACCTCTATGCAATCTGGGTACCCACT 287  
|||||  
326 uSer 327  
|||||  
286 GTCT 283  
seq\_name: gb\_est1:AL522373  
seq\_documentation\_block:  
LOCUS AL522373 960 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL522373 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CS0DB008YW02 3  
prime, mRNA sequence.  
ACCESSION AL522373  
VERSION AL522373.1 GI:12785866  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 960)  
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
source  
1. 960  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DB008YW02"  
/clone\_lib="LTI\_NFL004\_NBC2"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"

/note="Organ: brain; Vector: pcMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: [liang@lifetech.com](mailto:liang@lifetech.com) URL: <http://fulllength.invitrogen.com>"

BASE COUNT 225 a 259 c 255 g 217 t 4 others  
ORIGIN

alignment\_scores:  
Quality: 1179.50 Length: 231  
Ratio: 5.151 Gaps: 2  
Percent Similarity: 99.134 Percent Identity: 98.268

alignment\_block:  
US-09-357-675C-21 x AL522373/rev ..

Align seg 1/1 to reverse of: AL522373 from: 1 to: 960

```

98 ThrLeuHisLeuSerGluProLeuGlyLysLeuLeu.GluGluTyrT 114
|||||
959 ACCGTACACCTGCTCTGAACCACTGGTGAAC...TTTGAAGAATACA 913
|||||
114 hrgInLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHis 130
|||||
912 CCCAGCTTCGACGGGAATGTGGACTGTGGCTGCTGGGGGTTTCAT 863
|||||
131 GluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisVa 147
|||||
862 GAGCGTCCCAAGCTGGAGCACTCAGAAATCTACAAATTTCACGT 813
|||||
147 LLeuLeuAsnSerIysGlyAlaValAlaThrTyrArgLysThrHisL 164
|||||
812 GCTGCTGAACAGCAAGAGGGCAGTAGGCCACTTACAGGAACACATC 763
|||||
164 eucyAspValGluIleProGlyGlnGlyPrometCysGluSerAsnSer 180
|||||
762 TGTGTGACGTAGAGATTCCAGGCGAGGCGCTATGTGTGAAGCACTCT 713
|||||
181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyL 197
|||||
712 ACCATGCTGGGCCCACTTGTGACTACCTGTGACACACACAGGCA 663
|||||
197 sileGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeu 214
|||||
662 GATTGGTCTAGCTGTCTGTATGACATGCCG.TTCCCTGAACCTCTCTCG 614
|||||
214 LaleuAlaGlnAlaGlyAlaGluIleLeuThrTrpProSerAlaPheGly 230
|||||
613 CATTGGCTCAAGCTGGAGCAGAGATTACTTACTTCCCTTACGCTTTTGA 564
|||||
231 SerIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAla 247
|||||
563 TCCATTACAGGCCCAAGCCCACTGGAGGTGTGTGCGGGCGCTGTAT 514
|||||
247 eGluThrGlnCysTyrValAlaAlaAlaGlnCysGlyArgHisHisG 264
|||||
513 CGAAGCCAGTGTATGTAGTGCAGCAGCAGTGTGAGCAGCCACCATG 464
|||||
264 LlysArgAlaSerTyrGlyHisSerMetValAlaAspProTrpGlyThr 280
|||||
463 AGAAGAGAGAAATTATGGCCACAGCATGTGTGAGACCCCTGGGGA 414
|||||
281 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAs 297
|||||
413 GTGTGGCCCGCTGCTGTGAGGGGCCAGGCTGTGCTTGGCCCAATAGA 364
|||||
297 pleuAsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHis 314

```

```

|||||
363 CCTCACTACTCTGGACAGATTGGCCGACACACTGCTGTCTTCAGACCC 314
|||||
314 TGTATGProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
313 GCAGGCTTGACCTTATGGCAATCTGGGTGACCCACACTGTCT 273

```

seq\_name: gb-est2:BI757823

seq\_documentation\_block:

LOCUS BI757823 946 bp mRNA linear EST 25-SEP-2001  
DEFINITION 603030404P1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:5200560 5',  
mRNA sequence.

ACCESSION BI757823  
VERSION BI757823.1 GI:15749401  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 946)  
NIH-MGC <http://mhc.nci.nih.gov/>.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

COMMENT Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

plate: LLAM1502 row: h column: 01

High quality sequence stop: 708.

Location/Qualifiers

1..946

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="5200560"

/clone\_lib="NIH\_MGC\_114"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pcMV-SPORT6; Site\_1: NotI;  
Site\_2: EcoRV (destroyed); RNA source anonymous pool of 6  
male brains, age range 23-27 yo. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.5 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 019. Note:  
this is a NIH\_MGC Library."

BASE COUNT 213 a 264 c 270 g 199 t

ORIGIN

alignment\_scores:

Quality: 1169.00 Length: 291  
Ratio: 4.395 Gaps: 7  
Percent Similarity: 91.409 Percent Identity: 83.849

alignment\_block:  
US-09-357-675C-21 x BI757823 ..

Align seg 1/1 to: BI757823 from: 1 to: 946

```

1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeu 17
|||||
69 ATGCTGGCTTCATCACACAGGCCCTCCACAGATTCTGTCTCTGTG 118
|||||
17 sProGlyLeuArgGlnProGluLeuSerValLeuCysAlaGlnProArgP 34
|||||
119 TCTGAGACTCCGGATTAACCTCACTCAGTACTTGTGCTCAGCCAGGC 168

```

seq_name:	gb_est2:BI755493
seq_documentation_block:	
LOCUS	BI755493 813 bp mRNA
DEFINITION	603023243P1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5197716 5'
ACCESSION	mRNA sequence.
VERSION	BI755493
KEYWORDS	BI755493.1 GI:15747071
	EST.

```

SOURCE          human.
ORGANISM        Homo sapiens
REFERENCE       Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        1 (bases 1 to 813)
TITLE          NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL        National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT        Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cgapsb@mail.nih.gov
                Tissue Procurement: Life Technologies, Inc.
                cDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLAM1495 row: a column: 13
                High quality sequence stop: 745.

FEATURES
  Source
    1..813
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:5197716"
    /clone_lib="NIH_MGC_114"
    /lab_host="DH10B"
    /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
    male brains, age range 23-27 yo. Library is oligo-dT
    primed and directionally cloned (EcoRV site is destroyed
    upon cloning). Average insert size 1.5 kb, insert size
    range 1-3 kb. Library is normalized and enriched for
    full-length clones and was constructed by C. Gruber
    (Inuitrogen). Research Genetics tracking code 019. Note
    this is a NIH-MGC Library."
    this is a NIH-MGC Library."

BASE COUNT      155 a 241 c 225 g 192 t
ORIGIN

alignment_scores:
  Quality: 1135.50      Length: 243
  Ratio: 4.916          Gaps: 2
  Percent Similarity: 95.062  Percent Identity: 92.593

alignment_block:
  US-09-357-675C-21 x B1755493 ..

Align seg 1/1 to: B1755493 from: 1 to: 813

1 MetLeuGlyPheIlePhrArpProPHISArGPheLeuSerIleuency 17
|||||
87 ATGCTGGGCTTCATCCACAGGCTCTCTCAAGATTCCTGCTTCGTG 136
|||||
17 sProGlyLeuArGIllePProGlnIleuSerValIleuCySAlaGlnProArp 34
|||||
137 TCCTGGACATCCGGATACCTCAACTCTCAGACTTGTGGCTCACCCAGCG 186
|||||
34 roArGAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuValAla 50
|||||
187 CCAGACCCATGGCTATCTCTCTCTCTCCCTCGCAACGTGCCCGGTGGCT 236
|||||
51 ValCyGlnValIThrSerThrProaPlySGlnGlnAsnPhetYsThrcy 67
|||||
237 GGTGTCCAGTAACTCGACGCCAGACCAACACAGACTTTAAACATG 286
|||||
67 salaGluLeuValArGIluaIAlaIaIaIaIaIaIaIaIaIaIaIaIa 84
|||||
287 TCGTGTAGCTGATTGAGAGGCTCCAGACATGGGTGCTGCTGCTTC 336
|||||
84 euProGluAlaPheAspPheIleAlaIaArgAspProAlaGluThrIleuHIS 100
|||||
337 TGGCTTAGGACATTTGACTTCAATTCAGCGGACCTCGACAGACGCTACAC 386
|||||

```

```

101 LeuSerGluProLeuGlyGlyLeuLeuGluGluTyrThrGlnLeuA1 117
|||||
387 CTGCTGAACCACTGGGTGGAAACTTTGGAAATACACCCAGCTTGC 436
|||||
117 aaargGluGlyGlyLeuTyrLeuSerLeuGlyGlyPheHisGluArgGly 134
|||||
437 CAGGGAATGTGACTGTGCTGCTGGTGGTGGTTCACAGACGTGGCC 486
|||||
134 InsprTPGJugInThrGlnLysIleTyrAsnGlyHisValLeuLeuAsn 150
|||||
487 AAGACTGGAGCAGACTCAGAAAATCTACAATTGCTACGTGCTGCTAAC 536
|||||
151 SerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeuGlyAspVa 167
|||||
537 AGCAGCGGGCAGTAGTGCCACTTACAGAGACACATCTGTGTGACGT 586
|||||
167 LgIuIleProGlyGlnGlyPrometGlyGlySerAsnSerThrMetProG 184
|||||
587 AGAGATTCCAGGGCAGGGGCGCTATGTGTGCAAGCAACTTACCATGCTG 636
|||||
184 lypProSerLeuGlnSerProValSerThrProAlaGlyLysIleGlyLeu 200
|||||
637 GGGCCAGCTTGTGAGTCACTGTCTACGCGCAGCAGCAGCAGCAG 685
|||||
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG1 217
|||||
686 GCGTGTGCTATGACGAGCGGTTCCCTGAACCTCTCTGTCGATGGCTCA 735
|||||
217 n. AlaglyValagluIleLeuThrTyrProSerAlaPhe..... 229
|||||
736 GACGCTGGACCGCAGATCTTACCTTACCTTACCTTACCTTACCTTAC 785
|||||
230 .....GlySerIleThrGlyPro 235
|||||
786 GCCCGCCACTGGAGCGGTGCTGGGGCTT 813
|||||

```

seq\_name: gb\_est2:BI711300

seq\_documentation\_block:

LOCUS BI711300 625 bp mRNA linear EST 19-SEP-2001  
 DEFINITION id66f12.x1 Human insulinoma Homo sapiens cDNA 3' similar to  
 TR:076091 076091 NITRILASE HOMOLOG 1.; mRNA sequence.

ACCESSION BI711300 GI:15686995

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 625)  
 Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
 Lemisha,I., Pearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
 Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistein,A.,  
 Schmitt,A., Theising,B., Rittler,E., Ronko,I., Bennett,J., Cardenas  
 ,M., Gibbons,M., McCann,R., Cole,R., Tsagarelisvili,R., Williams,T.,  
 Jackson,Y. and Bowers,Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)

TITLE JOURNAL  
 COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138

Tel: 617-495-1812  
 Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. J. Ferrer in vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
 University Genome Sequencing Center for information on obtaining a  
 clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
 Seq primer: 40up from Gibco  
 High quality sequence stop: 481.

FEATURES

Location/Qualifiers

```

source
1. 625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human insulinoma"
/tissue_type="Insulinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site_1:
xhoI; Site_2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
BASE COUNT 139 a 183 c 161 g 142 t
ORIGIN

alignment_scores:
Quality: 1115.00 Length: 208
Ratio: 5.361 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-357-675c-21 x BI711300/rev ..

Align seg 1/1 to reverse of: BI711300 from: 1 to: 625

76 ArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleal 92
|||||
625 AGACTGGGTCCTGCGCTGCTTCCCTGCGCAGAGCATTTACTTCAATGCG 576
|||||
92 aargaspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysL 109
|||||
575 ACGGAGCCCTGCAGAGACGCTACACCTGTCTGAACCGCTGGTGGGAAC 526
|||||
109 euLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTyrPLeuSer 125
|||||
525 TTTTGGAAAGATACCCACCTTGCAGGGAATGTGACTGTGCTGTGCTGC 476
|||||
126 LeuGlyGlyPheHisGluArgGlyGlnAspTyrPGLuGlnThrGlnLysI 142
|||||
475 TTGGGTGGTTCATGATGAGCGGTGGCCAAAGACTGGGAGAGACCTCAAAAT 426
|||||
142 eTyrAsnGlyHisValLeuLeuAsnSerLysGlyAlaValAlaIleThrT 159
|||||
425 CTACAATTGTTCACGTGCTGCTGAACAGCAAGGGGCGAGTGTGGCCACTT 376
|||||
159 yTyrGlySerThrHisLeuGlyAspValGluIleProGlyGlnGlyPromet 175
|||||
375 ACAGGAAGACACATCTGTGTGACGTTCAGAGATTCAGAGGCGGGCTATG 326
|||||
176 CysGluSerAsnSerThrMetProGlyProSerLeuGlnSerProValise 192
|||||
325 TGTGAAGCAACCTTACCATGCGTGGGCCAGCTTGTGAGTCACCTGTGACG 276
|||||
192 rThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheP 209
|||||
275 CACACAGCAGCAGCAGATGTGTCTAGCTGTCTGTGATGACATCGGTGCC 226
|||||
209 rGluLeuSerLeuAlaLeuAlaGluAlaGluAlaGluIleLeuThrTyr 225
|||||
225 CTAACTCTCTGTGGCATTTGGCTCAAGCTGGAGCAGAGACTTACTTACTAT 176
|||||
226 ProSerAlaPheGlySerIleThrGlyProAlaHisTyrPLeuValleuLe 242
|||||
175 CTTTCAGCTTTTGGATTCATTACAGGCCACGCCAGCGGAGGTGTGCT 126
|||||
242 uArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaIleAGlnc 259
|||||

```

```
125 GCGGGCCCGTGTATCGAAACCCAGTGCCTATGATAGTGGCAGCAGCAGT 76
259 ysglyArgHISHisGluysArgAlaSerTyrglyHisSerMetValVal 275
75 GTGGACGCCACCATGAGAGAGAGCAAGTATAGGCCACAGCATGTGTGTA 26
276 AspProTrpGlyThrValValAla 283
25 GACCCCTGGGGACACAGTGTGGCC 2
```

**THIS PAGE BLANK (USPTO)**

OM of: US-09-357-675C-21 to: Issued\_Patents\_NA.\* out\_format : pfs  
Date: Apr 29, 2002 10:02 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL-frame+P2n.model -DEV-x1h  
-Q=/cgn2\_1/USPTO.spool/US09357675/turnat\_29042002\_091757\_3210/app\_query.fasta\_1.1208  
-DB-Issued\_Patents\_NA -OPMT=fastlap -SUFFIX=p2n.rn1  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -RGAPOP=6.000 -RGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blotsum2 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09357675\_@CGN1.118 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPRY -WAIT -THREADS=1

## Search information block:

Query: US-09-357-675C-21  
Query length: 327  
Database: Issued\_Patents\_NA.\*  
Database sequences: 38353  
Database length: 122816752  
Search time (sec): 85.140000

## Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-63 +		206.00	364.19	1.3e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-63 +		205.00	364.19	1.3e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-25 +		205.00	362.25	1.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-25 +		205.00	362.25	1.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-5 +		204.00	360.31	2.2e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-59 +		204.00	360.31	2.2e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-5 +		204.00	360.31	2.2e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-59 +		204.00	360.31	2.2e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-61 +		203.00	358.37	2.8e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-61 +		203.00	358.37	2.8e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-9 +		202.00	356.43	3.6e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-11 +		202.00	356.43	3.6e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-11 +		202.00	356.43	3.6e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-917-111-1 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-479-638-1 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-1 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-3 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-7 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-13 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-15 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-17 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-19 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-23 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-27 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-35 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-51 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-55 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-65 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-1 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-3 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-7 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-13 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-15 +		201.00	354.49	4.7e-12	1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-17 +		201.00	354.49	4.7e-12	1785

seq\_name: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-63

## seq\_documentation\_block:

Sequence 63, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Nanda, Hitokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1785 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
ORIGINAL SOURCE:  
STRAIN: JM109 PAD469  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 233..1144  
US-08-294-871A-63

alignment\_scores:  
Quality: 206.00 Length: 340  
Ratio: 1.280 Gaps: 15  
Percent Similarity: 47.353 Percent Identity: 25.294

## alignment\_block:

US-09-357-675C-21 x US-08-294-871A-63 ..

Align seg 1/1 to: US-08-294-871A-63 from: 1 to: 1785

```
50 AlavalCysGlnValThr.....SerThrProAs 59
   |||:|||||:|||||:|||||:|||||:|||||:
164 GCATGTGTTTCACAAAGCTTTCCGCGCGCTGGCCGACATCAGCTAG 213
   |||:|||||:|||||:|||||:|||||:|||||:
59 pLysGlnGlnAspNheLysThr..... 66
   |||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTTCATGACACGTGATGATCTTGCAGTGGGACAC 263
   67 .....Cys 67
264 AAGGTCCGATCGCGCGCGAGACACGACGACGAGTCTGCTGCTT 313
68 AlagluLeuValargluAlaAlaargluAlaCysLeuAlaPheLe 84
   |||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATGCTGACGAAGCCGCGAGCCGGCGCAATTTCATTGCTT 363
84 uproGlu.....AlaPheAspPheLeAlaArg.....Asp 95
   |||:|||||:|||||:|||||:|||||:|||||:
364 CCCGAAACTCGCGCTTACGACCTTCTCCGCGCTGCGCATTCACGACG 413
95 roAlagluThrLeuHisLeuSerGlu...ProLeuGluGlyLysLeuLeu 110
   |||:|||||:|||||:|||||:|||||:|||||:
414 AGCGCGAGCTCGATAGCTCTATGAGACCGAATAAGCCCGCGGTGCTC 463
111 GluGluTyrThrGlnLeuAlaargluCysGlyLeuThrLeuSerLeuG 127
   |||:|||||:|||||:|||||:|||||:|||||:
464 CGTCCACTCTTTGAGAAGCCGCGGAACTCGGATCGGCTTCAATCGGG 513
127 ygluPheHis.....GluargGlyGlnAspTrpGluInThrGlnL 141
   |||:|||||:|||||:|||||:|||||:|||||:
514 CTACGCTGAACCTGCTGCGAAGCGGC.....GTCAAGC 548
141 yslleTyraSncysHisValLeuLeuAsnSerLysGlyAlaValAla 157
   |||:|||||:|||||:|||||:|||||:|||||:
549 GTGCGCTTCAACAGCGCTATTGCTGATATAGTACGCAAGATCGTCGC 598
158 ThrTyArgLysThrHisLeuCysAspValGluIleProGlyGlnGly 174
   |||:|||||:|||||:|||||:|||||:|||||:
```

```
599 AAGTATCGTAAGATTCATTTC.....CCGGGTCAACAAG.. 631
174 omecYsgIuserAsnSerThrMetProGlyProSerLeuGluSerPro 191
   |||:|||||:|||||:|||||:|||||:|||||:
632 .....GAGTACGAGGCTTACCGCGCGTTCACGACATCTTGAAAGGCT 674
191 alSerThrPro.....AlaGly 196
   |||:|||||:|||||:|||||:|||||:|||||:
675 ATTTGAGCCGGCGATCTCGGCTCCCGCTATGACGTGACGCGCGCG 724
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213
   |||:|||||:|||||:|||||:|||||:|||||:
725 AAATGGGAGATTTCATCTGCAACGATCGCGCTGCTGCTGCTGCTGCT 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 230
   |||:|||||:|||||:|||||:|||||:|||||:
775 GGTGATGGCTTCAGGGCGCGCGGATCATCTGC.....G 809
230 LysSerIleThrGlyProAlaHisTrpGluValLeu..... 241
   |||:|||||:|||||:|||||:|||||:|||||:
810 GCGGCTACACAGCCGCGACATCCATGTTGTTCCCGCAGCAGCACAC 859
242 .....LeuargAlaargAlaIleGluTh 249
860 CTGACGCTCTCCACCATCTCTATGATGACAGCGCGGCTTATGAA 909
249 rgIncysTyrValValAlaAlaAlaIncysGlyArgHisGluLysA 266
   |||:|||||:|||||:|||||:|||||:|||||:
910 CGGGGCTGCTGCGCGCGCGCGGCAAGTGGGCTATG...GAGGAGACT 956
266 rGlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVal 282
   |||:|||||:|||||:|||||:|||||:|||||:
957 GCATGCTGCTGCGCGCATCTGATCTGCGCGCGCGGGAAGTCTG 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
   |||:|||||:|||||:|||||:|||||:|||||:
1007 GCTCTACTACAGACCTGTGAAGACGAGTGATCACCGCCCGCTGATCT 1056
298 uAsnTyrLeuArgGlnLeuValArgArgHisLeuProValPhe...GlnHis 314
   |||:|||||:|||||:|||||:|||||:|||||:
1057 CGATCGCTGCGGGAAGTGGTGAACACATCTTCACTTCAACAGCAGATC 1106
314 rGArgProAspLeuTyrGly 320
   |||:|||||:|||||:|||||:|||||:|||||:
1107 GTCAGCCCGCAGCATATGCT 1126
seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-8716-398A-63
seq_documentation_block:
; Sequence 63, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiko
; APPLICANT: NAMBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```



```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 PAD469
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-63

```

```

alignment_scores:
Quality: 206.00 Length: 340
Ratio: 1.280 Gaps: 15
Percent Similarity: 47.353 Percent Identity: 25.294

```

alignment\_block:

US-09-357-675C-21 x US-08-876-398A-63 ..

Align seg 1/1 to: US-08-876-398A-63 from: 1 to: 1785

```

50 AAlaValcysGlnValThr.....SerThrProAs 59
||||:|||||:|||||:|||||:|||||:|||||:
164 GCGATGTTGTTTACACACGTTTCCCGCGCGCTGGCGCGACATCACCCTAG 213
||||:|||||:|||||:|||||:|||||:|||||:
59 pLysGlnGlnAspPheLysThr..... 66
||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGACGAGGTTTCATGACACGTCAGATGATCTTGCATGGGACAC 263
67 .....Cys 67
264 AAGGTCCGATGCGCGCGGAGACGCGAAGGTCGTCGTCCTT 313
68 AAlaGluLeuValArgGlnAlaAlaArgLeuLysLysLeuAlaPheLe 84
||||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATGCTGACGAAGCCGAGCGCGGCGGAATTCATTGCTCTT 363
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95
|||||:|||||:|||||:|||||:|||||:|||||:
364 CCCGAGACTGGCGCTTAGACCTTCTTCCCGCGCTGGCATTTCCACCGAG 413
95 roAlaGluThrLeuHisLeuSerGlu..ProLeuGlyGlyLysLeuLeu 110
|||||:|||||:|||||:|||||:|||||:|||||:
414 AGGCGGAGCTGATAGCTTTATGAGACCGGAATGCCGCGGCTGCTC 463

```

```

111 GluGluTyrThrGlnLeuAlaArgGlnCysGlyLeuThrPheSerLeu 127
...:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
464 GCTCCACTCTTTGAGAGAGCCGCGGAACTCGGGATCGGCTTAAATCTGGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpGlnThrGlnL 141
|:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
514 CTACGCTGACACCGTCGTCGAAGCGGC.....GTCAGC 548
141 ySLeTyrAsnGlyHisValLeuLeuAsnSerLysGlyAlaValAla 157
..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
549 GTCCGCTCAACACGTCATTTTGTGATTAAGTCAGCAAGATCGTCGGC 598
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGly 174
|||||:|||||:|||||:|||||:|||||:|||||:
599 AAGTATCGTAAGATCCATTG.....CCGGTCACAAG.. 631
174 oMetCysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191
||:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
632 .....GAGTACGAGCGCTACCGCCGCTCCAGATCTTGAAGCGTT 674
191 aSerThrPro.....AlaGly 196
||:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
675 ATTTCGACCCGGCGATCTCGGCTTCCCGTCTATGACGTGACCGCCGG 724
197 LysIleGlyLeuAlaValAlcysTyrAspMetArgPheProGlnLeuSer 213
||||:|||||:..:..:..:..:..:..:..:..:..:..:..:..:..:
725 AAAATGGGATGTTCTGTCACAGATCGCCGCTGCGCTGAAGCCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 230
..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
775 GGTGATGGGCGCCACAGGGCGCGGACATCATCTGC.....G 809
230 LysSerIleThrGlyProAlaHisTrpGluValLeu..... 241
||||:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
810 GCGGCTAACACACGCGGACCCACAATCCCATTTGTCGCCAGCAGCAC 859
242 .....LeuArgAlaArgAlaIleGluThr 249
860 CTGACGCTCTTCCACCATCTCTATGATGACGCGCGGCTTATCAGAA 909
249 rGlnCysTyrValAlaAlaAlaAlaGlnCysGlyArgHisIleGluTyr 266
||||:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
910 CCGGGCTGTGTCGGCGCGCGCGGCGGCAAGGTGGCATG..GAGGAGACT 956
266 rGAlaSerTyrGlyHisSerMetValAlaAspProTrpGlyThrValAl 282
|||||:|||||:..:..:..:..:..:..:..:..:..:..:..:..:..:
957 GCATGCTGCTCGCCACATCTGTCATGCGCGCGGCGGGAATTCGTC 1006
283 AAlaArgCysSerGlu..GlyProGlyLeuCysLeuAlaArgIleAsp 298
||:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
1007 GCTTCACCTACGACGCTGGAGACGAGGTATCACCAGCGCGCTGATCT 1056
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
||:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
1057 CGATGCTGCGCGGAGAACTGCGTGAACACATCTTCAACTCAAGCAGCATC 1106
314 rGArgProAspLeuTyrGly 320
||:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:
1107 GTCAGCCCGCAGCATATGCT 1126

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-294-871A-25

seq\_documentation\_block:

; Sequence 25, Application US/08294871A

; Patent No. 5824522

; GENERAL INFORMATION:

; APPLICANT: Ikenaka, Yasuhiro

; APPLICANT: Namba, Hirokazu

; APPLICANT: Takano, Masayuki

; APPLICANT: Yajima, Kazuyoshi

; APPLICANT: Yamada, Yukio

; APPLICANT: Takahashi, Satomi

; APPLICANT: Okubo, Kazuma

APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraiishi, Yoshio  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1785 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
ORIGINAL SOURCE:  
STRAIN: JM109 PAD445

FEATURE:  
NAME/KEY: CDS  
LOCATION: 233..1144  
US-08-294-871A-25  
alignment\_scores:  
Quality: 205.00 Length: 335  
Ratio: 1.273 Gaps: 15  
Percent Similarity: 48.060 Percent Identity: 25.970  
alignment\_block:  
US-09-357-675C-21 x US-08-294-871A-25 ..  
Align seg 1/1 to: US-08-294-871A-25 from: 1 to: 1785  
50 AlaValGlyGlnValThr.....SerThrProAs 59  
164 GCGATGCTTTTCACACGCTTTCCCGCGCGCTGGCGGACATCACCCTAG 213  
59 PLYSGInGlnAsnPhelysthr..... 66  
214 GAAGAGACGAGGTTTCATGACACGTCAGATGATCTTCACATGGGACAC 263  
67 .....Cys 67  
264 AAGTCCGATCGCGCGCGGAGACACGGAACAGTGTCTGTCGTCCTT 313  
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84  
314 CTCGACATGTCGACGAACCGCGAGCGCGGCGGCAATTCATTCATTCCTT 363  
84 uProGlu.....AlaPheAspPheAlaArg.....Asp 95  
364 CCCGGAACCTCCGCTTACGACCTTCTCCCGCGCTGATTCACCGACG 413  
95 roAlaGluThrLeuHisLeuSerGlu..ProLeuGlyGlyLeuLeu 110  
414 AGCGCGAGCTCGATAGGCTTCTATGAGACGGAATCCCGCGCGTGTGTC 463  
111 GluGluThrThrGlnLeuAlaArgGluCysGlyLeuThrPheLeuSerLeu 127  
464 GTCGCACTTTTGAAAGCGCGCGAAGCTCGGATCGGCTTCATTCATGCG 513  
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141  
514 CTACGCTGAACCTCGTCGGAAGCGCG.....GTCAAGC 548  
141 ySLeuTyraAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157  
549 GTCCCTTCAACACGCTCCATTTTGGTGGATAGTCAGGCAGATGCTGCGC 598  
158 ThrTyraArgLysThrHisLeuCysAspValGluLeuProGlyGlnGly 174  
599 AAGTATCTAAGATCCATTG.....CCGGGTCAACAG.. 631  
174 omeCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191  
632 .....GAGTACGAGCGCTACCGCGCTTCACATCTTGAAGAAGCGTT 674  
191 alSerThrPro.....AlaGly 196  
675 ATTTCGAGCGCGGCGATCTCGGCTTCCCGGTCTATGACGTCGACCGCG 724  
197 LysIleGlyLeuAlaValCysTyraPheLarGheProGluLeuSerLe 213  
725 AAATGGGAGTTCATCTGCAACGATCGCGCTGCGCTGAAGCGTGGCG 774  
213 uAlaLeuAlaGlnAlaGlyAlaGluLeuThr.....TyrProSerA 228  
775 GGTATGCGCTCAGGGCGCGGAGATCATCTCGGCGGCTACACACACG 824  
228 lapheGlySerIleThrGlyProAlaHisTrpGluValLeu..... 241

```

      825 CGACCCACATCCACCGTCCCGACACGACCGCTGAGCTCTCTCCAC 874
      242 .....LeuArgAlaArgAlaIleGluThrGlnCysTyrValva 254
      875 CATCTCTTATGATGACGCGCGGTCTTATCAGAACGGGCGCTGTCGCG 924
      254 IAlaAlaAlaGlnCysGlyArgHisGlnGlyValArgAlaSerTyrGln 271
      925 GGGCGCGGGAAGGTGGCATG...GAGGAGAACTGCATGCTGCTCGCC 971
      271 IserMetValValAspProTyrGlyThrValValAlaArgCysSerGlu 287
      972 ACTCTCTCATCTGCGCGCGCGGGAATCGCTCTCATCAGACG 1021
      288 ...GlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgG1 303
      1022 CTGGAAAGACGAGTGATCACCGCGCGCTGATCTGATCTGCTCGGGA 1071
      303 nLeuArgArgHisLeuProValPhe...GlnHisArgArgProAspLeu 319
      1072 ACTGCGGAAACATCTTCAACTTCAACAGACATCTCAGCCCGACACT 1121
      319 YrGly 320
      1122 ATGGT 1126

```

seq\_name: /cgn2.6/ptodata/2/1na/6A.COMB.seq:US-08-876-398A-25

seq\_documentation\_block:

Sequence 25, Application US/08876398A

Patent No. 6083752

GENERAL INFORMATION:

APPLICANT: IKENAKA, Yasuhiro

APPLICANT: NAMBA, Hirokazu

APPLICANT: TAKANO, Masayuki

APPLICANT: YAJIMA, Kazuyoshi

APPLICANT: YAMADA, Yukio

APPLICANT: TAKAHASHI, Satomi

TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESS: FOLEY & LARDNER

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/876,398A

FILING DATE: 16-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/211,641

FILING DATE: 11-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/JP93/01101

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 340078/1992

FILING DATE: 21-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 212692/1992

FILING DATE: 10-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wegner, Harold C.

REGISTRATION NUMBER: 25,258

```

REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 PAD445
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-25

```

```

alignment_scores:
Quality: 205.00 Length: 335
Ratio: 1.273 Gaps: 15
Percent Similarity: 48.060 Percent Identity: 25.970

```

alignment\_block:

US-09-357-675c-21 x US-08-876-398A-25 ..

Align seg 1/1 to: US-08-876-398A-25 From: 1 to: 1785

```

50 AlAlaLysGlnValThr.....SerThrProAs 59
||||:||||:||||:||||:||||:||||:||||:
164 GCGATGTGTTTCACACAGTTTTCGCCGCGTGGCGGACATCACCTAG 213
59 PLSGInGlnAsnPhelThr..... 66
||||:||||:||||:||||:||||:||||:
214 GAAGAGCAGAGGTTTCATGACACGTCAGATGATTCTGACATGGGACAAC 263
67 .....Cys 67
264 AAGTCGATGCGCGCGCGGACACGCAACAGATGTCGTGCTT 313
68 AlAGluValArgGluAlaIleArgLeuGlyAlaCysLeuAlaPhe 84
||||:||||:||||:||||:||||:||||:
314 CTCGACATGTGACGAAAGCCGCGAGCGGGCGGAAATTCATTGCTCT 363
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95
||||:||||:||||:||||:||||:||||:
364 CCCGAACTCGCGCTTACGACCTTCTCCGCGCTGGCATTCACCGACG 413
95 roAlAGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuLeu 110
||||:||||:||||:||||:||||:||||:
414 AGCGCGACCTGATAGCTTATGAGACCGAAATCCGCGCGGTGTC 463
111 GlnGluTyrThrGlnLeuAlaArgGluCysGlyLeuTyrPheSerLeu 127
||||:||||:||||:||||:||||:||||:
464 CTCGACCTTTGAGAAAGCCGGAACCTCGGAGCTTCAATTCGCG 513
127 yGlyPheHis.....GluArgGlyGlnAspTyrPheGlnThrGln 141
||||:||||:||||:||||:||||:||||:
514 CTACGCTGACCTCGTCGTCGAAAGCGGC.....GTCAGGC 548
141 ySLeTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
||||:||||:||||:||||:||||:||||:
549 GTCGCTTCACACGTCATTTTGTGGATAGTCAGCAAGATCGTCGCG 598
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGly 174
||||:||||:||||:||||:||||:||||:
599 AAGTATCGTAAGATCCATTG.....CCGCGGCACAG.. 631
174 oMetCysLeuSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191
||||:||||:||||:||||:||||:||||:
632 .....GAGTACGAGGCGCTACCGCGCTTCACAGATCTTGAAGAAGCGTT 674

```

```

191 alserThrPro.....AlaGly 196
      |||
675 ATTCGAGCGGCGCATCTCGGCTCCCGGTATGATGCTGACGCGCGG 724
      |||
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerIle 213
      |||
725 AAATGGGGATGTTTCATCTCAACGATCCGCCGCTGAGCCTGAGCGG 774
      |||
213 uAlaLeuAlaGlnAlaGlyAlaGluLeuThr.....TyrProSerA 228
      |||
775 GGTGATGGGCTCAGGGCGCCGAGATCATCTCGCGGCTTACACACGCG 824
      |||
228 laphGlySerIleThrGlyProAlaHisTrpGluValLeu..... 241
      |||
825 CGACCCACATCCACCGCTCCCGACAGACACACCGAGCTCCTTCAC 874
      |||
242 .....LeuArgAlaArgAlaIleGluThrGlnCysTyrValVa 254
      |||
875 CATCTCTATCGATCGAGCGCGGCTTATCAGAACGCGGCGCTGTCGCG 924
      |||
254 lAlaAlaIleGlnCysGlyArgHisGluLysArgAlaSerTyrGlyH 271
      |||
925 GGCCTGGGGAGAGTGGGCTG...GAGGAGAACTGCATGCTGCTGCGGC 971
      |||
271 lSerMetValValAspProTrpGlyThrValAlaAlaArgCysSerGlu 287
      |||
972 ACTCTCGCATCGTGGCGCGCGGGAATCGCTGCTCCTACTACGACG 1021
      |||
288 ...GlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgI 303
      |||
1022 CTGAGAGACGAGATGATCACCGCGCGCTGATCTGATCGTGGCGGA 1071
      |||
303 nLeuArgArgHisLeuProValPhe...GlnHisArgArgProAspLeu 319
      |||
1072 ACTGCGGAGACATCTCAACTCAAGACGATCGTCAACCCGACGACT 1121
      |||
319 YrGly 320
      |||
1122 ATGCT 1126

```

seq.name: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-5

seq.documentation\_block:

Sequence 5, Application US/08294871A

Patent No. 5824522

GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro

APPLICANT: Namba, Hirokazu

APPLICANT: Takano, Masayuki

APPLICANT: Yajima, Kazuyoshi

APPLICANT: Yamada, Yukio

APPLICANT: Takahashi, Satomi

APPLICANT: Okubo, Kazuma

APPLICANT: Yamada, Kazuhiko

APPLICANT: Hiraishi, Yoshiro

TITLE OF INVENTION: Immobilized Enzyme Preparation and

TITLE OF INVENTION: Process for Producing D-a-Amino Acid

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/294,871A

```

FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JM109, PAD404 (FERM BP-3913)
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-5

```

alignment\_scores:

Quality	Ratio	Length
204.00	1.267	340
Percent Similarity: 47.353	Percent Identity: 25.294	Gaps: 15

alignment\_block:

US-09-357-675C-21 x US-08-294-871A-5 ..

Align seg 1/1 to: US-08-294-871A-5 from: 1 to: 1785

```

50 AlayaIcysGlnValThr.....SerThrProAs 59
|||:||||:||||:||||:||||:||||:||||:
164 GCATGTGTTTCACACGTTTCCCGCGCGCTGGCGCGGACATCACCTAG 213

```

59 pLysGlnGlnAenPheLysThr..... 66  
|||:|||||:|||||  
214 GAGGAGCAGAGGTTTCATGACACGTCAGTGTACTTGCAGTGGGACAC 263  
67 .....Cys 67  
264 AAGTCCGATCCGCGCGGAGACACGCGAAGAGTGGTGGTCTT 313  
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84  
:|||||:|||||:|||||  
314 CTCGACATGCTGACGAAAGCCGCGACCCGCGCAATTTCATTCCTT 363  
84 uProGlu.....AlaPheAspPheLeuAlaArg.....Asp 95  
:|||||:|||||:|||||  
364 CCCCAGACCTCGCGCTTACGACCTTCTCCCGCGCTGGCATTTCCACCGAG 413  
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyLeuLeu 110  
|||||:|||||:|||||  
414 AGCGGAGCTCGATAGCTTCTATGAGACCGAATGCCCGCGGTGGTC 463  
111 GluGluTyrThrGlnLeuAlaArgGlyCysGlyLeuTrpLeuSerLeu 127  
:|||||:|||||:|||||  
464 CGTCCACTCTTGGAGAGCGCGCGAAGCTCGGGATCGCTCAATCTGG 513  
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141  
|||:|||||:|||||  
514 CTACGCTGACTCGTCTGAGAGCGGC.....GTCAGC 548  
141 ySileTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157  
:|||||:|||||:|||||  
549 GTCGCTTCAACACGTCATTTGGTGATAGTCAGGCAAGATCGTGGC 598  
158 ThrTyrArgLysThrHisLeuCysAspValGluLeuProGlyGlnGly 174  
|||||:|||||:|||||  
599 AAGTATCGTAAGATCCATTTC.....CCGGTCACAAAG... 631  
174 oMetCysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191  
|||:|||||:|||||  
632 .....GAGTGCAGAGCCTACCGCGCTTCCAGCATCTTGAAGACGTT 674  
191 aSerThrPro.....AlaGly 196  
|||:|||||:|||||  
675 ATTTCAGCGCGCGCATTCGCGTCCCGGTATGACGTCAGCGCGG 724  
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213  
|||||:|||||:|||||  
725 AAAATGGGATGTCATCTGCAACGATCCGCGCTGAGAGCCTGGCG 774  
213 uAlaLeuAlaGlnAlaGlyAlaGluLeuThrTyrProSerAlaPhe 230  
:|||||:|||||:|||||  
775 GGTGATGGCTCAGGCGCGGAGTCAATCTGC..... 809  
230 LysSerIleThrGlyProAlaHisTrpGluValLeu..... 241  
|||:|||||:|||||  
810 GCGGCTACAAACACGCGGACCCCAATCCCTGTGTCCAGACGACAC 859  
242 .....LeuAlaGlnAlaArgAlaIleGluThr 249  
:|||||:|||||:|||||  
860 CTGAGCTCTTCACCATCTCTATGATGACAGCGGCTTATCAGAA 909  
249 rGlnCysTyrValValAlaAlaGlnCysGlyArgHisHisGluLys 266  
:|||||:|||||:|||||  
910 GCGGCTGGTCCGCGCGCGGCAAGTGGGCTATG...GAGGAAACT 956  
266 rGlnAspSerGlyHisSerMetValAlaAspTrpGlyThrValAla 282  
|||||:|||||:|||||  
957 GCATGCTGCTGGCCACTCTGATCTGGCGCGGCAAGTGGTCTG 1006  
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298  
|||:|||||:|||||  
1007 GCTCTCATTACGACGCTGGAGACGAGGTATCACCGCGCGCTGATCT 1056

298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314  
|||:|||||:|||||  
1057 CGATCGCTGCGGGAAGTCTGCTGACACATCTTCACTTCAACGACATC 1106  
314 rGArgProAspLeuTyrGly 320  
|||:|||||:|||||  
1107 GTCAGCCCGCAGCACTATGCT 1126  
seq\_name: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:US-08-294-871A-59  
seq\_documentation\_block:  
; Sequence 59, Application US/08294871A  
; Patent No. 5824522  
; GENERAL INFORMATION:  
; APPLICANT: Ikenaka, Yasuhiro  
; APPLICANT: Namba, Hirokazu  
; APPLICANT: Takano, Masayuki  
; APPLICANT: Yajima, Kazuyoshi  
; APPLICANT: Yamada, Yukio  
; APPLICANT: Takahashi, Satomi  
; APPLICANT: Okubo, Kazuma  
; APPLICANT: Yamada, Kazuhiko  
; APPLICANT: Hiraiishi, Yoshiro  
; TITLE OF INVENTION: Immobilized Enzyme Preparation and  
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/294, 871A  
; FILING DATE: 22-AUG-1994  
; CLASSIFICATION: 435  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/971,758  
; FILING DATE: 12-APR-1993  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/917,111  
; FILING DATE: 07-AUG-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 08/211,641  
; FILING DATE: 11-APR-1994  
; APPLICATION DATA:  
; APPLICATION NUMBER: JP 400848/1990  
; FILING DATE: 07-DEC-1990  
; APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP91/01696  
; FILING DATE: 06-DEC-1991  
; APPLICATION DATA:  
; APPLICATION NUMBER: JP 407922/1990  
; FILING DATE: 27-DEC-1990  
; APPLICATION DATA:  
; APPLICATION NUMBER: JP 078840/1991  
; FILING DATE: 11-APR-1991  
; APPLICATION DATA:  
; APPLICATION NUMBER: JP 140051/1991  
; FILING DATE: 12-JUN-1991  
; APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP92/00739  
; FILING DATE: 10-JUN-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: JP 212692/1992  
; FILING DATE: 10-AUG-1992  
; APPLICATION DATA:

```

APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JM109 PAD456
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-59

```

```

alignment_scores:
Quality: 204.00      Length: 340
Ratio: 1.267        Gaps: 15
Percent Similarity: 47.353      Percent Identity: 25.294

```

```

alignment_block:
US-09-357-675C-21 x US-08-294-871A-59 ..

```

```

Align seg 1/1 to: US-08-294-871A-59 from: 1 to: 1785

```

```

50 AlAlaValCysGlnValThr.....SerThrProAs 59
|||:::|||||:::|||||:::|||||:::|||||:::
164 GCGATGCTGTTTACACACAGTTTCCGGCGCGTGGCGGACATCACCAG 213
|||:::|||||:::|||||:::|||||:::|||||:::
59 pLysGlnGlnAsnRheLysThr..... 66
|||:::|||||:::|||||:::|||||:::|||||:::
214 GAAGGAGCAGAGTTTCATGACACAGTCAGATGATCTTCAGTGGGACAAC 263
|||:::|||||:::|||||:::|||||:::|||||:::
67 .....Cys 67
264 AAGTCCGATGCGCGCGCGGAGACACCGCAACAGTGTGCTGCTT 313
68 AlAGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
|||:::|||||:::|||||:::|||||:::|||||:::
314 CTCGACATGCTGACGAAAGCCGCGGCGCGGAGTTTCATGCTGCTT 363
|||:::|||||:::|||||:::|||||:::|||||:::
84 uProGlu.....AlaPheAspRheLeuAlaArg.....Asp 95
|||:::|||||:::|||||:::|||||:::|||||:::
364 CCCCAGCTCCGCTTACGACCTTCTCCCGCGCTGATTTTACCAGGAG 413
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuSerGlu 110
|||:::|||||:::|||||:::|||||:::|||||:::
414 AGCCCGAGCTGATAGCTTTATGAGACCGAAATCCCGGCGCGGTGGTC 463
|||:::|||||:::|||||:::|||||:::|||||:::
111 GluGluThrGlnLeuAlaArgGlnCysGlyLeuThrLeuSerLeuGln 127
|||:::|||||:::|||||:::|||||:::|||||:::
464 CGTCCACATCTTTGAGAGCGCGGAGACTCGGAGATCGGCTTCAATCTGG 513
|||:::|||||:::|||||:::|||||:::|||||:::
127 yGlyPheHis.....GluArgGlyGlnAspThrLeuGlnThrGlnL 141
|||:::|||||:::|||||:::|||||:::|||||:::
514 CTACGCTGAACCTGCTGTCGAAGCGGCG.....GTCAAGC 548
|||:::|||||:::|||||:::|||||:::|||||:::
141 ySileThrAspCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
|||:::|||||:::|||||:::|||||:::|||||:::
549 GTCCGTCACACAGTCATTTTGGTAAAGTCAGCAAGATCGTGGC 598

```

```

158 ThrTyArgLysThrHisLeuCysAspValGluLeuProGlyGlnGlyP 174
|||:::|||||:::|||||:::|||||:::|||||:::
599 AAGATTCGTAAGATCCATTTG.....CCGGGTACACAAG... 631
|||:::|||||:::|||||:::|||||:::|||||:::
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
|||:::|||||:::|||||:::|||||:::|||||:::
632 .....GAGTACGAGGCTTACCGGCGCTTCAGCATCTTGAAGAGCTT 674
|||:::|||||:::|||||:::|||||:::|||||:::
191 alSerThrPro.....AlaGly 196
|||:::|||||:::|||||:::|||||:::|||||:::
675 ATTTCGAGCCGGCGCATCTCGGCTTCCCGGTATGAGCTGACGCCGCG 724
|||:::|||||:::|||||:::|||||:::|||||:::
197 LysIleGlyLeuAlaValCysTyArgPheArgPheProGluLeuSerLe 213
|||:::|||||:::|||||:::|||||:::|||||:::
725 AAATGGGATGTTTCATCTGCAAGCATGCGCGCTGCTGAAGCTGGCG 774
|||:::|||||:::|||||:::|||||:::|||||:::
213 uAlaLeuAlaGlnAlaGlyAlaGluThrTyThrProSerAlaPheG 230
|||:::|||||:::|||||:::|||||:::|||||:::
775 GGTGATGGGCTTCAGGGCGCGGAGATCATTCG.....G 809
|||:::|||||:::|||||:::|||||:::|||||:::
230 LysIleThrGlyProAlaHisThrPgluValLeu..... 241
|||:::|||||:::|||||:::|||||:::|||||:::
810 GCGGTACACACAGCGCGGACCCACATCCCTGTGTCCCGACAGCAGCAC 859
|||:::|||||:::|||||:::|||||:::|||||:::
242 .....LeuArgAlaArgAlaIleGluTh 249
|||:::|||||:::|||||:::|||||:::|||||:::
860 CTGACGCTCTTCCACCATCTCTATCGATGACGCGCGGTCTTATCAGA 909
|||:::|||||:::|||||:::|||||:::|||||:::
249 rGlnCysTyArgValAlaAlaAlaAlaGlnCysGlyArgHisGlnLys 266
|||:::|||||:::|||||:::|||||:::|||||:::
910 CGGGGCTGTGCGCGCGCGCGGCAAGTCAGGCAATG...GAGGAGACT 956
|||:::|||||:::|||||:::|||||:::|||||:::
266 rGAlaSerTyGlyHisSerMetValAlaAspProThrPgluThrVal 282
|||:::|||||:::|||||:::|||||:::|||||:::
957 GCATGCTGCTGCGCCATCTCTGATCGTGGCGCGGCAAGTATCGT 1006
|||:::|||||:::|||||:::|||||:::|||||:::
283 AlArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
|||:::|||||:::|||||:::|||||:::|||||:::
1007 GCTCTCATACGACGCTGGAAGACGAGTGATCACCGCGCGCTGAGCT 1056
|||:::|||||:::|||||:::|||||:::|||||:::
298 uAsnTyLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
|||:::|||||:::|||||:::|||||:::|||||:::
1057 CGATGCTGCGCGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106
|||:::|||||:::|||||:::|||||:::|||||:::
314 rGArgProAspLeuTyGly 320
|||:::|||||:::|||||:::|||||:::|||||:::
1107 CTCACGCCACGACATATGCT 1126
|||:::|||||:::|||||:::|||||:::|||||:::

```

```

seq_name: /cgn2.6/plodata/2/ina/6a_COMB.seq:US-08-876-398A-5

```

```

seq_documentation_block:

```

```

; Sequence 5, Application US/08876398A
; Patient No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKABA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 PAD404 (FERM BP-3913)
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-5
```

```
alignment_scores:
Quality: 204.00      Length: 340
Ratio: 1.267         Gaps: 15
Percent Similarity: 47.353   Percent Identity: 25.294
```

alignment\_block:  
US-09-357-675C-21 x US-08-876-398A-5 ..

Align seg 1/1 to: US-08-876-398A-5 from: 1 to: 1785

```
50 AlaValCysGlnValThr.....SerThrProAs 59
|||||:|||||:|||||:|||||:|||||:|||||:
164 CGCATGTGTTTCACACAGCTTTTCCGCGCGCTGGCCGACATCAGCTAG 213
59 pLysGlnGlnAsnPhelysthr..... 66
|||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGACGACAGAGGTTCATGACAGTCAGATGATCTTGACAGTGACAC 263
67 .....Cys 67
264 AAGTCCGATCCGCGCGCGGAGACACGACAGCTGCTGCTCTT 313
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhele 84
|||||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATCTGTGCGAAGCCCGGACCGCGGCGCAATTTCAATGCTT 363
84 uproGlu.....AlaPheAspPheleAlaArg.....Asp 95
|||||:|||||:|||||:|||||:|||||:|||||:
364 CCCGACACTCGCGCTTACGACCTTCTTCCGCGCTGCGCATTTCCACG 413
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuLeu 110
|||||:|||||:|||||:|||||:|||||:|||||:
```

```
414 AGCCGAGCTCATAGCTTCTATGACACGAAATGCCGCGCGCTGTC 463
111 GluGluThrThrGlnLeuAlaArgGluCysGlyLeuThrLeuSerLeuG 127
|||:|||||:|||||:|||||:|||||:|||||:
464 GGTCCACTCTTTGAGAGAGCGCGGAACTCGGATCGGTTCAATCTGG 513
127 yGluPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141
|||:|||||:|||||:|||||:|||||:|||||:
514 CTACGCTGAACTGCTGTCGAAAGCGGC.....GTCAGC 548
141 yLysLeuAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
|||:|||||:|||||:|||||:|||||:|||||:
549 GTCGCTTCAACAGCTCCATTTGTTGGATGAATGACGACAGATCTGCGC 598
158 ThrTyrArgLysThrHisLeuCysAspValGlnLeuProGlyGlnGly 174
|||||:|||||:|||||:|||||:|||||:|||||:
599 AAGATTCGTAAAGATCCATTG.....CCGGGTCAAG.. 631
174 ometCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
|||:|||||:|||||:|||||:|||||:|||||:
632 .....GAGTACGAGGCTTACCGCGCTTCCAGCATCTTGAAGAGCTT 674
191 alSerThrPro.....AlaGly 196
|||:|||||:|||||:|||||:|||||:|||||:
675 ATTTCGAGCCGCGGATCTGCGCTTCCCGGTATGACGTGACGCCGCG 724
197 LysLleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213
|||||:|||||:|||||:|||||:|||||:|||||:
725 AAATGGGATGTTCACTCATGCAACGATCGCGCGCTGAGAGCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluLeuThrTyrProSerAlaPheG 230
|||||:|||||:|||||:|||||:|||||:|||||:
775 GGTGATGGCTCAGGCGCGCGCGGATCATCTGC.....G 809
230 LysSerLethrGlyProAlaHisTrpGluValLeu..... 241
|||||:|||||:|||||:|||||:|||||:|||||:
810 GCGGCTCAACAGCGCGACCCACATCCCTTGTGCCAGACAGCACAC 859
242 .....LeuArgAlaArgAlaLleGluThr 249
|||||:|||||:|||||:|||||:|||||:|||||:
860 CTGACGCTCTCCACATCTCTATCGATCGACGCGCGCTTATTCAGAA 909
249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisGluLysA 266
|||:|||||:|||||:|||||:|||||:|||||:
910 GCGGCGCTGTCGCGCGCGCGGCAAGTGGCATG..GAGGAGACT 956
266 rGlnAserTyrGlnHisSerMetValValAspProTrpGlyThrVal 282
|||||:|||||:|||||:|||||:|||||:|||||:
957 GCATGCTGCTCGGCGCACTCTGCATCGTGGCGCGGAGGAATGCTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgLysPhe 298
|||:|||||:|||||:|||||:|||||:|||||:|||||:
1007 GCCTCTCACTACAGCGCTGGAAGACGAGTGATGATACCCGCCCTG 1056
298 uAsnTyrLeuAlaArgGlnLeuArgHisLeuProValPhe...GlnHis 314
|||||:|||||:|||||:|||||:|||||:|||||:
1057 CGATCGCTCGCGGAACTGCTGTAACATCTTCACTTCAAGGAGATC 1106
314 rGArgProAspLeuTyrGly 320
|||||:|||||:|||||:|||||:|||||:|||||:
1107 GTCAGCCCGACGACTATGCT 1126

seq_name: /cgn2.6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-59
seq_documentation_block:
; Sequence 59, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKANO, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
```

```

APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
TITLE OF INVENTION: 70
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 pad456
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
IS-08-876-398A-59

```

```

alignment_scores:      Quality: 204.00      Length: 340
                        Ratio:    1.267      Gaps:    15
Percent Similarity:    47.353      Percent Identity: 25.294

alignment_block:
US-09-357-675C-21 x US-08-876-398A-59      ..
Align seg 1/1 to: US-08-876-398A-59 from: 1 to: 1785

```

```

50  AAlaValCysGlnValThr.....SetThrproms 59
    |||:::|||||
164  GCcAtGtGtTTCACAAcGTTTCCCGGCCCTGGCGCGGACATCACTAG 213
    .....
59  pLysGlnGlnAsnDhelysthr..... 66
    |||:::|||||
214  GAAGGAGCCACAGGTCCTATGACACGTCAGATCATCTTCGATGGGACAC 263
    .....

```

```

67 ..... Cys 67
264 AAGTCCGATCGCGCGCGGAGACAGCGGACAGGTCGTCTGCTT 313
68 AlaGluLeuValaArgGlnAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
   ::::: ::::: ||||| ||||| :::::
314 CTCGACATGCTGACGAAAGCGCGGCGCGGCGGCGGAAATTTCATTGCTT 363
   ||||| .....AlaPheAspPheLeuAlaArg.....Asp 95
84 uProGlu.....AlaPheAspPheLeuAlaArg.....Asp 95
   ||||| ||||| ||||| |||||
364 CCGCGACACTCGCGTTACAGACTCTTCCGCGCTGGATTTCACGACG 413
95 roAlaGluPthLeuHisLeuSerGlu..ProLeuGlyLeuLeuLeu 110
   ||||| ::::: ||| ::::: ||||| |||||
414 AGCGCGAGCTCGATAGCTCTTATGATACCGAAATGCGCGCGCGGTGC 463
111 GluGluPthGlyThrGlnLeuAlaArgGluCysGlyLeuThrPheSerLeuG1 127
   ::::: ::::: ||| ||| ||||| ::::: |||||
464 CGTCCACTCTTTGAAAGAGCGCGGAGAACTCGGATCGGCTTACGTGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141
   ::::: ::::: ||| ||| |||||
514 CTAGCGTAAGCTCGTCTCGAAGCGCG.....GTCAAGC 548
141 ySLeuArgAsnCysHisValLeuLeuLeuAsnSerGlyAlaValAla 157
   ::::: ||||| ::::: ||||| ::::: |||||
549 GTCCGCTTCAACAGCTCCATTTTGTGGATGAAGTAGGACAGACATCGTGGC 598
158 ThrTyArgGlyThrHisLeuCysAspArgAlaGluLeuProGlyGlnGlyPr 174
   ||||| ||||| ||||| |||||
599 AAGATATCGTAAATGATCCATTTC.....CCGGTCCACAG.. 631
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
   ||| ::::: ||| |||||
632 .....GAGTACGAGGCGTCCGCGCGCTTCCACATCTTGAAAGCGTT 674
191 aSerThrPro.....AlaGly 196
   |||
675 ATTTCGAGCGCGGCGATCTCGGCTCCGCGTATGACGTGACGCGCGG 724
197 LysLeuGlyLeuAlaValaCysTyAspMetArgPheProGlyLeuSerLe 213
   ||||| ||||| ::::: ||| ||||| |||||
725 AAAATGGGAGATGTCATCTGCAACGATCGCGCGCTGTAAGCTTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluLeuThrTyProSerAlaPheG 230
   ::::: ||||| ||||| |||||
775 GGTGATGGGCGTCAAGGCGCGGAGATCATCTGC.....G 809
230 LysSerLeuThrGlyProAlaHisTrpLeuValLeu..... 241
   ||||| ::::: ||||| |||||
810 CGCGCTCAACACCGCGCACCCACATCCCTGTCTCCACAGCAGCAGCAG 859
242 .....LeuArgAlaArgAlaLeuGluTh 249
   ||||| ::::: ||| ::::: |||||
860 CTGACGTCTTCCACATCTCTTATGATGAGCGCGGCGTATTTCAGAA 909
249 rGlnCysTyValaValaAlaAlaGlnCysGlyArgHisHisGluCysA 266
   ::::: ::::: ||||| ||||| |||||
910 CGGGGCGCTGTCGCGCGCGCGGCGGCAAGTCAGCATG..GAGAGAACT 956
266 rGlnAlaSerTyGlyHisSerMetValaValaAspProTrpGlyThVala 282
   ||||| ||||| ||||| |||||
957 GCATGCTGCTCGCGCACTCTCGATCTGTCGGCGCGGCGGGAATGCTT 1006
283 AlaArgCysSerGlu..GlyProGlyLeuCysLeuAlaArgLeuAspLe 298
   ||| ::::: ||| ::::: |||||
1007 GCTCTCACTGACGCGTGGAAAGAGAGAGTGTACCGCGCGCGCTCTT 1056
298 uAsnTyLeuArgGlnLeuArgArgHisLeuProValPhe..GlnHisA 314
   ::::: ||||| ||||| ||||| |||||
1057 CGATTCGCTCGCGGAAGTGGGTGAACACATTTTCACTTCAAGAGCATC 1106
114 rGArgProAspLeuTyGly 320

```



seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-294-871A-61  
1107 GTCAGCCCGACGACTATGCT 1126  
||::||::||::||::||  
seq\_documentation\_block:  
; Sequence 61, Application US/08294871A  
; Patent No. 5824522  
; GENERAL INFORMATION:  
; APPLICANT: Ikenaka, Yasuhiro  
; APPLICANT: Nanda, Hirokazu  
; APPLICANT: Takano, Masayuki  
; APPLICANT: Yajima, Kazuyoshi  
; APPLICANT: Yamada, Yukio  
; APPLICANT: Takahashi, Satomi  
; APPLICANT: Okubo, Kazuma  
; APPLICANT: Yamada, Kazuhiko  
; APPLICANT: Hiraiishi, Yoshiro  
; TITLE OF INVENTION: Immobilized Enzyme Preparation and  
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/294,871A  
; FILING DATE: 22-AUG-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/971,758  
; FILING DATE: 12-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/917,111  
; FILING DATE: 07-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/211,641  
; FILING DATE: 11-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 400848/1990  
; FILING DATE: 07-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP91/01696  
; FILING DATE: 06-DEC-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 407922/1990  
; FILING DATE: 27-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 078840/1991  
; FILING DATE: 11-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 140051/1991  
; FILING DATE: 12-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP92/00739  
; FILING DATE: 10-JUN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213692/1992  
; FILING DATE: 10-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP93/01101  
; FILING DATE: 05-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 340078/1992  
; FILING DATE: 21-DEC-1992

ATTORNEY/AGENT INFORMATION:  
; NAME: Wegner, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 61:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1785 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; ORIGINAL SOURCE:  
; STRAIN: JMI09 PAD468  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 233..1144  
; US-08-294-871A-61  
  
alignment\_scores:  
Quality: 203.00 Length: 340  
Ratio: 1.261 Gaps: 15  
Percent Similarity: 47.353 Percent Identity: 25.294  
  
alignment\_block:  
US-09-357-675c-21 x US-08-294-871A-61 ..  
Align seg 1/1 to: US-08-294-871A-61 from: 1 to: 1785  
  
50 AlaValCysGlnValThr.....SerThrProAs 59  
|||::|||::|||::|||::|||::|||::|||::|||  
164 GCGATGTTGTTTACAAAGCGTTTCCGCGCGCGCGGACATCACCCTAG 213  
|||::|||::|||::|||::|||::|||::|||::|||  
59 PLYSGInGlnAspPheLysThr..... 66  
|||::|||::|||::|||::|||::|||::|||::|||  
214 GAAGGAGCAGAGGTTTCATGACACGTGATGATCTTGCATGGGAGAAC 263  
|||::|||::|||::|||::|||::|||::|||::|||  
67 .....Cys 67  
  
264 AAGTCGATCGCGCGCGCGGAGACAGCAGAGTGTGCTGCTT 313  
|||::|||::|||::|||::|||::|||::|||::|||  
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84  
|||::|||::|||::|||::|||::|||::|||::|||  
314 CTCGACATGCTGACGAAGCGCGGAGCGCGCGCATTTTCATTGCTT 363  
|||::|||::|||::|||::|||::|||::|||::|||  
84 UProGlu.....AlaPheAspPheIleAlaArg.....Asp 95  
|||::|||::|||::|||::|||::|||::|||::|||  
364 CCCCGAAGTCGCGCTTACGACCTTTCCCGCGCTGCGATTTCACCGACG 413  
|||::|||::|||::|||::|||::|||::|||::|||  
95 roAlaGluThrIleuHisLeuSerGlu...ProLeuGlyGlyLysLeu 110  
|||::|||::|||::|||::|||::|||::|||::|||  
414 AGCGCGAGCTCGATAGCTTCTATGAGACCGAATGCCCGCGCGGTGTC 463  
|||::|||::|||::|||::|||::|||::|||::|||  
111 GluGluItyrThrGlnLeuAlaArgGluCysGlyLeuThrLeuSerLeu 127  
|||::|||::|||::|||::|||::|||::|||::|||  
464 CGTCACACTCTTTGAGAAAGCGCGGAGACTGGGATCGCTTCAATCGGG 513  
|||::|||::|||::|||::|||::|||::|||::|||  
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141  
|||::|||::|||::|||::|||::|||::|||::|||  
514 CTCAGCTGAAGCTGCTGTCGAAGCGGC.....GTCAGC 548  
|||::|||::|||::|||::|||::|||::|||::|||  
141 ySileTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157  
|||::|||::|||::|||::|||::|||::|||::|||  
549 GTCGCTTCAACAGCTCCATTTTGTGATTAAGCAGCAAGATCGTCGCG 598  
|||::|||::|||::|||::|||::|||::|||::|||  
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyLysGln 174  
|||::|||::|||::|||::|||::|||::|||::|||  
599 AAGTATCGTAAAGATCCATTGG.....CCGGGTCACAAG.. 631  
|||::|||::|||::|||::|||::|||::|||::|||  
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191

```

632 .....GAGTACGAGGCTTACCGGCGTTCACGACATCTTGAAGCGCTT 674
191 alserThPro.....Alagly 196
675 ATTTGACCGGGCGATCTCGGCTTCCGGTCTATGACGTCGACCGCGG 724
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213
725 AAAATGGGATGTTTCATCTGCAACGATCGCGCTGGCCGGAAGCTGGG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 230
775 GGTGATGGGCTTCAGGGCGCCGACATCATCTGC.....G 809
230 LysSerIleThrGlyProAlaHisTyrPgluValLeu..... 241
810 CGGGCTACACACGCGGACCCACAAATCCCGCTGTTCCCGACGACGAC 859
242 .....LeuArgAlaArgAlaIleGlu 249
860 CTGACGCTCTCCACCATCTCTATGATGACGCGCGGCTTATCAGAA 909
249 rGlnCysTyrValValAlaAlaGlnCysGlyArgHisGluLysA 266
910 CGGGCTGTCGTCGCGCGCGGCGGCAAGTGGGATG...GAGGAGAACT 956
266 rGAlaSerTyrGlyHisSerMetValValAspProTyrPgluThrVal 282
957 GCATGCTCTGCGCCATCTGCAATCTGCGCGGCGGCAAGTGGGAAATG 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
1007 GCTCTCATACGACGCTGGAAGACGAGGATGACCGCGCGCTGATCT 1056
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
1057 CGATGCTGCGCGGGAACGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106
314 rArgProAspLeuTyrGly 320
1107 GTCAAGCCCGACCATATGCT 1126

seq_name: /cgn2.6/prodata/2/lna/6A_COMB.seq:us-08-876-398A-61

seq_documentation_block:
; Sequence 61, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayoshi
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 pad468
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-61

```

```

alignment_scores:
  quality: 203.00      length: 340
  ratio: 1.261          gaps: 15
  percent similarity: 47.353  percent identity: 25.294

```

```

alignment_block:
US-09-357-675c-21 x US-08-876-398A-61 ..

```

```

Align seg 1/1 to: US-08-876-398A-61 from: 1 to: 1785

```

```

50 AlaValCysGlnValThr.....SerThrProAs 59
||||:||||| :|||
164 GCGATGTTTCACAACGTTTCCGCGCGCTGGCGCGGACATCATCAG 213
59 pLysGlnGlnAsnPheLysThr..... 66
||||:||||| :|||
214 GAAGGAGCAGAGGTTTCATGACACGTCAGATGATCTGACGTGGGACAC 263
67 .....Cys 67
264 AAGTTCGATCCGCGCGGAGACACGCAAGAGTGTGCTGCTT 313
68 AlaGluLeuValArgLysAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
||||:||||| :|||
314 CTGACATGCTGACGAAGCCGCGAGCGGCGGGAATTTCATTGTCTT 363
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95
||||:||||| :|||
364 CCGGACATCGGCTTACGACCTTCTCCGCGCTGGCATTTCCACCGAG 413
95 rAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeu 110
||||:||||| :|||
414 AGCGGAGCTGATGATGCTTCTATGAGACGGAATGCGCGCGGTGTC 463
111 GluGluTyrThrGlnLeuAlaArgLysGlyLeuTyrPheSerLeu 127
||||:||||| :|||
464 GGTCACTCTTGTGAGAGGCCGCGGAACCTCGGATCGGCTTCATCTGGG 513

```

127 yglphenis.....GluArgIleGlnAspTrpGluInThrGlnL 141  
1 111 111  
514 CTACGGCTGAACTCGTCGAAAGCGGC.....GTCAGC 548  
141 ysIleTrpAsnGlyHisValLeuLeuAsnSerIleGlyAlaValAla 157  
1 111 111  
549 GTCGGCTTCAACACGTCATTTTGTGATTAATGACAGCAAGATCGCGC 598  
158 ThrTrpArgLysThrHisLeuGlyAspValGluIleProGlyGlnGly 174  
1 111 111  
599 AAGTATCGTAAGATCCATTG.....CCGGGTCACAAG.. 631  
174 oMetCysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191  
1 111 111  
632 .....GAGTACGAGCGCTACCGCGCTTCCAGCATCTTGAAACGCT 674  
191 aIleSerThrPro.....AlaGly 196  
1 111  
675 ATTTGAGCGCGGCGATCTCGGCTCCGGTCTATGACGTCAGCGCGCG 724  
197 LysIleGlyLeuAlaValCysTrpAspMetArgPheProGluLeuSer 213  
1 111 111  
725 AAAATGGGATGTTCATGTGCAACGATCGCGCTGAGCTGAGCTGGCG 774  
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTrpProSerAlaPhe 230  
1 111 111  
775 GGTGATGGGCGCTCAGGGCGCGCATCATCTGC.....G 809  
230 LysSerIleThrGlyProAlaHisTrpGluValLeu..... 241  
1 111 111  
810 GCGGCTACACACGCGCCGACCAATCCGCTGTCTCCCGACGACGAC 859  
242 .....LeuArgAlaArgAlaIleGluThr 249  
1 111 111  
860 CTGACGCTCTTCACCATCTCTATGATGAGGAGCGGGCTTATCAGAA 909  
249 rGlnCysTrpValValAlaAlaIleGlnCysGlyArgHisIleGlnLys 266  
1 111 111  
910 CGGGGCGCTGTCGGCGCGCGCGCAAGTGGGCATC..GAGAGACACT 956  
266 rAlaSerTrpGlyHisSerMetValValAspProTrpGlyThrVal 282  
1 111 111  
957 GCATGCTGCTCGGCGCATCTCTGATCGTGGCGCGACCGGGAATCGTC 1006  
283 AlaArgCysSerGlu...GlyProGlyLeuGlyLeuAlaArgIleAsp 298  
1 111 111  
1007 GCTCTCTACTACGACGCTGGAAGAAGAGGATGATCACCGCGCGTCT 1056  
298 uAsnTrpLeuArgGlnLeuArgHisLeuProValPhe...GlnHis 314  
1 111 111  
1057 CGATCGCTCCCGGAGACTGCGTAACATCTTCAACTTCAACGACGATC 1106  
314 rGArgProAspLeuTrpGly 320  
1 111 111  
1107 GTCAGCCCGCACCATATGCT 1126

seq\_name: /cgn2\_6/ptodata/2/ine/5A\_COMB.seq:US-08-294-871A-9

seq\_documentation\_block:  
: Sequence 9, Application US/08294871A  
: Patent No. 5824522  
: GENERAL INFORMATION:  
: APPLICANT: Ikenaka, Yasuhiro  
: APPLICANT: Nambu, Hirokazu  
: APPLICANT: Takano, Masayuki  
: APPLICANT: Yajima, Kazuyoshi  
: APPLICANT: Yamada, Yukio  
: APPLICANT: Takahashi, Satomi  
: APPLICANT: Okubo, Kazuma  
: APPLICANT: Yamada, Kazuhiko  
: APPLICANT: Hirata, Yoshiro  
: TITLE OF INVENTION: Immobilized Enzyme Preparation and

TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AORA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1785 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
ORIGINAL SOURCE:  
STRAIN: JM109 PAD416 (FERM BP-3915)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 233..1144

US-08-294-871A-9

Alignment\_scores:  
Quality: 202.00 Length: 340  
Ratio: 1.255 Gaps: 15  
Percent Similarity: 47.353 Percent Identity: 25.294

Alignment\_block:  
US-09-357-675c-21 x US-08-294-871A-9 ..

Align seg 1/1 to: US-08-294-871A-9 from: 1 to: 1785

```
50 AlaValcysGlnValThr.....SerThrProAs 59
|||:|||||:|||||:|||||:|||||:|||||:
164 GCGATGTGTTTCACAAACGTTTCCCGCGCGTGGCGGACATACACCTAG 213
|||:|||||:|||||:|||||:|||||:|||||:
59 pLysGlnGlnAsnPhelLysThr..... 66
|||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTTCATGACACGTCAGATGATCTTCAGTGGGACAA 263
.....Cys 67
264 AAGTCCGATCGCGCGCGGAGACACGCCAAGAGTCTGCTGCTCTT 313
68 AlaGlnLeuValArgLysAlaAlaArgLeuGlnAlaCysLeuAlaPheLe 84
|||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATGCTGACGAAGCGCGGACCGCGGCGCAATTTCATCTCTT 363
84 uProGlu.....AlaPheAspPheLeAlaArg.....Asp 95
|||:|||||:|||||:|||||:|||||:|||||:
364 CCGACGAACCTCGCGCTTACGACCTCTCTCCGCGCTGGCATTTACCGAC 413
95 roAlaGlnThrLeuHisLeuSerGlu..ProLeuGlnGlyLysLeuLeu 110
|||:|||||:|||||:|||||:|||||:|||||:
414 AAGCCGAGCTGATAGCTTCTATGAGACCGGAATGCCCGCGCGGAGTC 463
111 GlnGluThrThrGlnLeuAlaArgLysGlnGlyLeuThrPheLeuSerLeu 127
|||:|||||:|||||:|||||:|||||:|||||:
464 CGTCACTCTTTGAGAGAGCGCGGACACTCGGATCGGCTTCATCTGGG 513
127 yGlyPheHis.....GluArgGlnGlnAspTrpGlnGlnThrGln 141
|||:|||||:|||||:|||||:|||||:|||||:
514 CTACGCTGAACTCGTCTGCAAGGCGGC.....GTCAAAC 548
141 ySileThrAsnGlyHisValLeuLeuAsnSerLysGlnAlaValAla 157
|||:|||||:|||||:|||||:|||||:|||||:
549 GTCCGCTTCAACACGTCATTTGGTGATAGTCAGGCAAGATCTCGGC 598
158 ThrTyrrArgLysThrHisLeuGlnCysAspValGlnLeuPheLeu 174
|||:|||||:|||||:|||||:|||||:|||||:
599 AAGTATCGTAAGATCCATTG.....CCGGTCACAAAG... 631
174 cMetCysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191
|||:|||||:|||||:|||||:|||||:|||||:
632 .....GAGTACGAGGCGCTACCGCGCTTCCAGCATCTTGAAGAAGCG 674
191 aLserThrPro.....AlaGly 196
|||:|||||:|||||:|||||:|||||:|||||:
675 ATTTCGAGCGCGGCGATCTCGGCTTCCCGGTCTATGACGTCAGCGCGG 724
197 LysIleGlyLeuAlaValCysTyrrAspMetArgPheProGlnLeuSerie 213
|||:|||||:|||||:|||||:|||||:|||||:
725 AAAATGGGATGTTCATCTGCAAGATCGCGCGCTGGAAGCCTCGGCG 774
213 uAlaLeuAlaGlnAlaGlnAlaGlnLeuThrTyrrProSerAlaPheG 230
|||:|||||:|||||:|||||:|||||:|||||:
775 GGTGATGGGCGCTCAGGCGCGGAGATCATCTGC.....G 809
230 LysSerIleThrGlyProAla..... 236
|||:|||||:|||||:|||||:|||||:|||||:
810 GCGGCTACACACGCCGACCAATCCCTGTTCCCGACGACGACAC 859
```

```
237 .....HisTrpGlnValLeuLeuArgAlaArgAlaIleGluTh 249
|||:|||||:|||||:|||||:|||||:|||||:
860 CTGACGCTCTCCACCATCTCTATGATGACGAGCGCGCTTATGAGAA 909
249 rGlnCysTyrrValAlaAlaAlaGlnGlnCysGlyArgHisHisGlnLys 266
|||:|||||:|||||:|||||:|||||:|||||:
910 CGGGGCTGTGCGCGCGCGCGGACGAGGCGGATG...GAGGAGAACT 956
266 rGlnSerTyrrGlnHisSerMetValAlaAspProTrpGlnThrValAl 282
|||:|||||:|||||:|||||:|||||:|||||:
957 GCATGCTGTGCGCGCATCTCTGATGCTGCGCGGACGCGGGAATCTGC 1006
283 AlaArgCysSerGlu..GlyProGlyLeuCysLeuAlaArgIleAspLe 298
|||:|||||:|||||:|||||:|||||:|||||:
1007 GCTCTCAGTACGACGCTGGAAGACGAGGTGATCACCGCGCGCTGATCT 1056
298 uAsnTyrrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
|||:|||||:|||||:|||||:|||||:|||||:
1057 CGATCGCTGCGCGGAACTGCTGAAACACATCTTCAACTTCAAGACGATC 1106
314 rArgProAspLeuTyrrGly 320
|||:|||||:|||||:|||||:|||||:|||||:
1107 CTCAGCGCCACGACATGCT 1126

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-294-871A-11
seq_documentation_block:
: Sequence 11, Application US/08294871A
: Patent No. 5824522
: GENERAL INFORMATION:
: APPLICANT: Ikenaka, Yasuhiro
: APPLICANT: Nambu, Hirokazu
: APPLICANT: Takano, Masayuki
: APPLICANT: Yajima, Kazuyoshi
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
: APPLICANT: Yamada, Kazuhiko
: APPLICANT: Hiraiishi, Yoshiro
: TITLE OF INVENTION: Immobilized Enzyme Preparation and
: NUMBER OF SEQUENCES: 70
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/294,871A
: FILING DATE: 22-AUG-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,758
: FILING DATE: 12-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/917,111
: FILING DATE: 07-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/211,641
: FILING DATE: 11-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 400848/1990
: FILING DATE: 07-DEC-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP91/01696
: FILING DATE: 06-DEC-1991
```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JM109 PAD428
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-11

alignment_scores:
Quality: 202.00 Length: 340
Ratio: 1.255 Gaps: 15
Percent Similarity: 47.353 Percent Identity: 25.294

alignment_block:
US-09-357-675C-21 x US-08-294-871A-11 ..
Align seg 1/1 to: US-08-294-871A-11 from: 1 to: 1785

50 AlalaGysGlnValThr.....SerThrProAs 59
|||||:|||||:|||||:|||||:|||||:|||||:
164 GCGAGTGTGTTTCAACAGCTTTCCCGCGCGCTGGCGGACATCAGCTAG 213
59 pLysGlnGlnAsnPhelYsThr..... 66
|||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTTCATGACAGTCAGATGATCTTGACAGTGGACAAC 263
67 .....Cys 67

264 AAGGTCCGATCGCGCGCGGAGACAGCGAAGAGGTGCTGCTGCTTT 313
68 AlagIuLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
|||||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATCTGACGAAGCCGCGAGCCGGCGCAATTTCTATTTGCTT 363
84 uProGlu.....AlaPheAspPheLeuAlaArg.....Asp 95
|||||:|||||:|||||:|||||:|||||:|||||:
364 CCCCAGAACTCGCGCTTACGACCTTCTTCCCGCGCTGCAATTCACGAG 413
```

```

95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuLeu 110
|||||:|||||:|||||:|||||:|||||:|||||:
414 AGGCCAGCTCCGATAGCTTCTATGACAGCGAAATGCCCGCGGTRGTC 463
111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 127
|||||:|||||:|||||:|||||:|||||:|||||:
464 CGTCACTCTTTGAGAAGCGCGGAACTCGGATCGGCTTCAATCGGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpGlnThrGln 141
|||||:|||||:|||||:|||||:|||||:|||||:
514 CTACGGCTGAACCTGCTGTCGAGAGCGCG.....GTCAGC 548
141 ySLeTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
|||||:|||||:|||||:|||||:|||||:|||||:
549 GTGCTTCAACAGCTGCATTTTGTGATAGTCAGGCAAGATGCTGCG 598
158 ThrTyrArgLysThrHisLeuCysAspValGluLeuProGlyGln 174
|||||:|||||:|||||:|||||:|||||:|||||:
599 AAGTATCGTAAATCCATTG.....CCGGTCAAG.. 631
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
|||||:|||||:|||||:|||||:|||||:|||||:
632 .....GAGTACGAGCGCTACCGCGCTTCAGCATCTGAAAGCGTT 674
191 alSerThrPro.....AlaGly 196
|||||:|||||:|||||:|||||:|||||:|||||:
675 ATTTCCAGCCGGCGATCTCGGCTCCCGCTATGACGTGACGCCGCG 724
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213
|||||:|||||:|||||:|||||:|||||:|||||:
725 AAATGGGATGTTTCATCTGCAACAGATCCGCGCTGAGCTGAGCGTGG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluLeuThrTyrProSerAlaPhe 230
|||||:|||||:|||||:|||||:|||||:|||||:
775 GGTGATGGCGCTCAGGGCGCGCAGATCACTGC.....G 809
230 lYserIleThrGlyProAla..... 236
|||||:|||||:|||||:|||||:|||||:|||||:
810 GCGGCTACACAGCCGACCCACAAATCCCTGTTCCCGACAGCAC 859
237 .....HisTrpGluValLeuLeuArgAlaArgAlaIleGluThr 249
|||||:|||||:|||||:|||||:|||||:|||||:
860 CTGACGCTCTCCACCATCTCTATGATGACAGCGCGGCTTATACGAA 909
249 rGlnCysTyrValAlaAlaAlaGlnCysGlyArgHisGluLysA 266
|||||:|||||:|||||:|||||:|||||:|||||:
910 CGGCGCTGTGTCGCGCGCGGCAAGCGCGCATG...GAGGAGAACT 956
266 rGlnAspSerTyrGlnHisSerMetValAlaAspProTrpGlyThrVal 282
|||||:|||||:|||||:|||||:|||||:|||||:
957 GCATGCTGCTGGCCACTCTGATCGTGGCGCGACCGGGAATGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
|||||:|||||:|||||:|||||:|||||:|||||:
1007 GCCTCCTACATCAGACGCTGGAAGAGAGGTATGATACCGCGCGGATCT 1056
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
|||||:|||||:|||||:|||||:|||||:|||||:
1057 CGATCGCTCGCGGGAAGCTGCTGAGACATCTTCAACTTCAAGAGCATC 1106
314 rGlnArgProAspLeuTyrGly 320
|||||:|||||:|||||:|||||:|||||:|||||:
1107 GTACGCCCAAGCACTATGCT 1126

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-294-871A-21
seq_documentation_block:
; Sequence 21, Application US/08294871A
; Patent No. 582432
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Namba, Hirokazu
; APPLICANT: Takano, Masayuki
```

APPLICANT: Yajima, Kazuyoshi  
 APPLICANT: Yamada, Yukio  
 APPLICANT: Takahashi, Satomi  
 APPLICANT: Okubo, Kazuma  
 APPLICANT: Yamada, Kazuhiko  
 APPLICANT: Hiraiishi, Yoshio  
 TITLE OF INVENTION: Immobilized Enzyme Preparation and  
 NUMBER OF SEQUENCES: 70  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/294,871A  
 FILING DATE: 22-AUG-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/971,758  
 FILING DATE: 12-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/917,111  
 FILING DATE: 07-AUG-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/211,641  
 FILING DATE: 11-APR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 400848/1990  
 FILING DATE: 07-DEC-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/JP91/01696  
 FILING DATE: 06-DEC-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 407922/1990  
 FILING DATE: 27-DEC-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 078840/1991  
 FILING DATE: 11-APR-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 140051/1991  
 FILING DATE: 12-JUN-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/JP92/00739  
 FILING DATE: 10-JUN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 212692/1992  
 FILING DATE: 10-AUG-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/JP93/01101  
 FILING DATE: 05-AUG-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 340078/1992  
 FILING DATE: 21-DEC-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Wegner, Harold C.  
 REGISTRATION NUMBER: 25,258  
 REFERENCE/DOCKET NUMBER: 74129/127/NOPA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 672-5300  
 TELEFAX: (202) 672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1785 base pairs  
 TYPE: nucleic acid

STRANDEDNESS: double  
 TOPOLOGY: linear  
 ORIGINAL SOURCE:  
 STRAIN: JM109 PAD439  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 233..1144  
 US-08-294-871A-21  
 alignment\_scores:  
 Quality: 202.00 Length: 340  
 Ratio: 1.255 Gaps: 15  
 Percent Similarity: 47.353 Percent Identity: 25.294  
 alignment\_block:  
 US-09-357-675C-21 x US-08-294-871A-21 ..  
 Align seg 1/1 to: US-08-294-871A-21 from: 1 to: 1785  
 50 AlaValCysGlnValThr.....SerThrProAs 59  
 |||:||||| :|||  
 164 GCGATGTGTTCCACAAAGTTTCCGGCCGCGCGGACATCACTG 213  
 59 pLysGlnGlnAsnPhelYsThr..... 66  
 |||:||||| :|||  
 214 GAAGGAGCAGAGTTTCATGACACGTCAGATGATCTTGCAGTGGACAAAC 263  
 67 .....Cys 67  
 264 AAGTCCGATCGCGCGCGGAGACACGCGAAGTCTGCTGCTT 313  
 68 AlaGlnLeuValArgGlnAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84  
 :|||:||||| :|||  
 314 CTCGACATCTGACGAAGCCGCGGAGCGCGGCGGAATTCATGTCCT 363  
 84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95  
 :|||:||||| :|||  
 364 CCCGAACTCGCGCTTACGACCTTCTCCGCGCTGGCATTTCCACGAG 413  
 95 roLalGluThrLeuHisLeuSerGlu...ProLeuGlyLeuYsLeuAla 110  
 |||:||||| :|||  
 414 AGCGGAGCTCGATAGCTTCTATGAGACCGAATGCGCGCGGATGTC 463  
 111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 127  
 :|||:||||| :|||  
 464 CCTCAGCTTTGAGAGAGCGCGGAACTCGGATCGCTCAATCGG 513  
 127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141  
 :|||:||||| :|||  
 514 CTACGCTGAACCTCGTCGCAAGCGGC.....GTCAAAC 548  
 141 yLleTyrAsnCysHisValLeuLeuAsnSerGlyAlaValAla 157  
 :|||:||||| :|||  
 549 GTCCGCTTCAACAGTCATTTGTGGATGATGAGCAAGATGTCGCG 598  
 158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnLeu 174  
 :|||:||||| :|||  
 599 AAGTATCGTAATCATCTTG.....CCGGTCACAAAG... 631  
 174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191  
 :|||:||||| :|||  
 632 .....GAGTACGAGGCTTACCGCGCTTCCAGCATTTGAAAAGCGTT 674  
 191 alSerThrPro.....AlaGly 196  
 :|||:||||| :|||  
 675 ATTTCGAGCGGCGCATCTCGGCTTCCGCTATGACGTCACGCGCG 724  
 197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213  
 :|||:||||| :|||  
 725 AAATGGGATGTTTCATCTGCAACGATCGCGCTGCGCTGAAGCGCGG 774  
 213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 230

```
775 GGGATGGGCTCAGGGGGCGGAGATCATCTCC.....G 809
230 YSerIleThrGlyProAla..... 236
810 GGGGTACACACAGCCGACCAATCCCCGTGTCCCGACGACGAC 859
237 .....HisTrpGluValLeuArgAlaArgAlaIleGluThr 249
860 CTGACGCTCTTCACCATCTCTATGATGACAGCGCGGTCTATCAGA 909
249 rGlnCysTrpValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
910 CGGGGCGCTGTCGGCGCGCGGCAAGGTGCGATG...GAGGAAAC 956
266 rGAlaSerTrpGlyHisSerMetValAlaAspTrpGlyThrValAl 282
957 GCATGCTGCTCGCCACCTCTGATGTCGCGCGGACCGGGAATCGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
1007 GCTCTACACACGACGCTGACAGACGAGTGATCACCAGCGCGCTCAT 1056
298 uAsnTrpLeuArgGlnLeuArgHisLeuProValPhe...GlnHisA 314
1057 CGATCGCTGCGGGAGACTGGTGACACACATCTTCACATTCAACGACATC 1106
314 rGArgProAspLeuTrpGly 320
1107 GTCAGCCCGGACGACATATGAT 1126
```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-9

```
seq_documentation_block:
; Sequence 9, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
```

```
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 PAD416 (FERM BP-3915)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-9

alignment_scores:
Quality: 202.00 Length: 340
Ratio: 1.255 Gaps: 15
Percent Similarity: 47.353 Percent Identity: 25.294

alignment_block:
US-09-357-675c-21 x US-08-876-398A-9 ..
Align seg 1/1 to: US-08-876-398A-9 from: 1 to: 1785

50 AlValCysGlnValThr.....SerThrProAs 59
164 GCATGTGTTCACACAGCTTTCCGCGCGCTGGCGGACATCACCTAG 213
59 pLysGlnGlnAsnPhelLysThr..... 66
214 GAAGAGACAGAGGTTCTATGACAGTCAGATGATACATTGACATGGGACAC 263
67 .....Cys 67
264 AAGTCCGATCGCGCGCGGACACGCGACAGTCGTCGTCCTT 313
68 AlGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
314 CTGACATGCTGACGAAAGCGCGGAGCGCGGCGGATTCATTGCTT 363
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95
364 CCCGAACTCGCGCTTACGACCTTCTCCGCGCTGACATTTCACGACG 413
95 rAlaGluThrIleHisLeuSerGlu...ProLeuGlyLysLeuLeu 110
414 AGCCCGAGCTCGATAGCTTATGAGACCGAAATCCCGCGGTGTC 463
111 GluGluTrpThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuG 127
464 GTCGCACTCTTGAGAAAGCGCGGAACTGGGATCGGCTTCATATCGGG 513
127 yGlyPheHis.....GluArgGlyLysAspTrpGluGlnThrGln 141
514 CTACGCTGACCTGCTCGTGAAGCGCGC.....GTCNAGC 548
141 ySleTrpAsnGlyHisValLeuLeuAsnSerLysGlyAlaValAla 157
549 GTCGCTTCAACAGTCCTCATTTGTGGATGATGTCAGCAAGATCGTCGG 598
158 ThrTrpArgLysThrHisLeuCysAspValGluIleProGlyGlnGly 174
599 AAGTATGTAAGATCATTTG.....CCGGTCACAAG.. 631
```





```
464 CGTCCACTCTTTGAGAAGCGCGGAACTCGGAGCTCGCTTCAATCTGGG 513
127 yGIpPheHis.....GLuArgGLyGlnAspTIpGLuGlnThrGlnL 141
514 CTACGGCTGAAGCTGCTCGTCGAGCGGC.....GTCAGC 548
141 ysIleTyraNcysHisValleuLeuAsnSerLysGlnAlaValAla 157
549 GTGCGCTTCAACAGCTCCATTGTTGGTGATAGTCAGGCAAGATCGTCGCC 598
158 ThrTyraGlySerThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
599 AAGTATCGTAAGATCCATTG.....CCGGTCCACAG.. 631
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
632 .....GAGTCAGAGCGCTACCGCGCTTCCAGCATCTTGAAGACGCTT 674
191 aIserThrPro.....AlaGly 196
675 ATTTCCAGCGCGGCGATCTGGCTCCCGGCTATGACGTGACGCCGCG 724
197 LysIleGlyLeuAlaValCysTyraSpmetArgPheProGluLeuSerLe 213
725 AAATGGGGGATGTTTCATCTGCAAGCATGCGCGCTGAGCTGAGCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyraProSerAlaPheG 230
775 GGTGATGGCGCTCAGGGCGCGCATCTGC.....G 809
230 LySerIleThrGlyProAla..... 236
810 GCGGCTACACACGCGACCCACACATCCCTGTTCGCCAGCAGACCAC 859
237 .....HisTyrGluValLeuLeuArgAlaArgAlaIleGluThr 249
860 CTGACGTCCTTCACCATCTCTATGATGATGACGCGCGCTTATCAGAA 909
249 rGlnCysTyraValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
910 CCGGGCCTGGTCCGCGCGCGGCAAGCGCGCATG..GAGGAGAACT 956
266 rGlnAserTyraGlyHisSerMetValAlaAspProTyrGlyThrVal 282
957 GCATGCTGCTCGGCACCTCTGCACTGGCGCGGACCGGGGAAATCGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
1007 GCTCTCACTACGACGCTGGAAGAGAGTATCATCCGCCCGCTGATCT 1056
298 uAsnTyraLeuArgGlnLeuArgHisLeuProValPhe...GlnHisA 314
1057 CGATCGCTGCGGGAACTCGGTGAACATCTTCAACTCAAGCAGCATC 1106
314 rGArgProAspLeuTyraGly 320
1107 GTGAGCCCCAGCACTATGTT 1126
```

**THIS PAGE BLANK (USPTO)**

OM of: US-09-357-675C-22 to: GenBank: \* out\_format : pfs

Date: Apr 29, 2002 11:10 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

#### Command line parameters:

```
-MODEL=frame+p2n.model -DEV=x1h  
-o=/gen2.1/USFO/spool/US09357675/runat_29042002_091757_3179/app-query.fasta_1.1208  
-db=genbank1 -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELPO=6.000 -DELPEXT=7.000 -START=1 -MATRIX=blomsu62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09357675 @CGN1_1 10065 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPX -WAIT -THREADS=1
```

#### Search information block:

Query: US-09-357-675C-22  
Query length: 323  
Database: GenBank: \*  
Database sequences: 1797656  
Database length: 1873333701  
Search time (Sec): 3374.980000

#### score list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation
gb_ro:AF069988	+ 1731.00	2473.72	2.0e-129	1338	AF069988 Mus musculus nitrilase
gb_ro:BC021634	+ 1723.00	2462.72	8.9e-129	1365	BC021634 Mus musculus, nitrilase
gb_pr:AF069987	+ 1487.00	2123.28	6.6e-110	1385	AF069987 Homo sapiens nitrilase
gb_ro:AF069985	+ 1347.00	1912.15	3.8e-98	4481	AF069985 Mus musculus nitrilase
gb_hlg:AC084821	+ 1343.00	1872.83	5.9e-96	211771	AC084821 Mus musculus chrom
gb_hlg:AC087229	+ 1320.00	1840.49	3.7e-94	213203	AC087229 Mus musculus chrom
gb_pr:AF069984	+ 1211.50	1718.52	2.2e-87	4079	AF069984 Homo sapiens nitrilase
gb_hlg:AL354714.4	+ 1211.50	1689.83	9.2e-86	110000	Continuation (5 of 7) of AL3
gb_hlg:AL354714.5	+ 1211.50	1689.83	9.2e-86	110000	Continuation (6 of 7) of AL3
gb_hlg:AL590651	+ 1211.50	1684.59	1.5e-85	167863	AL590651 Homo sapiens chrom
gb_pr:AL591806	+ 1211.50	1684.59	1.5e-85	200822	AL591806 Homo DNA sequence
gb_hlg:AC105589	+ 1187.00	1650.46	1.4e-83	178418	AC105589 Rattus norvegicus c
gb_ov:AF284575	+ 933.50	1330.14	1.0e-65	1214	AF284575 Xenopus laevis NIT pr
gb_in:AF069986	+ 706.00	1002.53	1.8e-47	1385	AF069986 Caenorhabditis elegans
gb_hlg:AC069989	+ 699.50	992.38	6.5e-47	1521	AF069989 Drosophila melanogaster
gb_hlg:AC093121	+ 699.50	953.53	9.5e-45	131853	AC017755 Drosophila melanogaster
gb_in:AC093501	+ 699.50	951.40	1.2e-44	168250	AC093501 Drosophila melanogaster
gb_hlg:AC006169	+ 699.50	951.10	1.3e-44	174279	AC006169 Drosophila melanogaster
gb_in:AC105293	+ 699.50	950.33	1.4e-44	190351	AC105293 Drosophila melanogaster
gb_in:AE003457	+ 699.50	946.41	2.4e-44	298640	AE003457 Drosophila melanogaster
gb_pl:AF372904	+ 635.50	902.81	6.3e-42	1172	AF372904 Arabidopsis thaliana
gb_hlg:AC010013	+ 624.00	856.98	2.3e-39	34013	AC010013 Drosophila melanogaster
gb_ba:AE000934	+ 538.50	741.34	6.3e-33	15124	AE000934 Methanobacterium th
gb_pl:SPBC651	+ 529.00	723.21	6.4e-32	25350	AL033570 S.pombe chromosome I
gb_pl:SCY0116W	+ 524.00	741.86	5.9e-33	1307	AF284574 Homo sapiens NIT prote
gb_pr:AF284574	+ 483.00	685.62	7.9e-30	962	AF284574 Homo sapiens NIT prote
gb_pr:AF260334	+ 483.00	685.19	8.4e-30	1019	AF260334 Homo sapiens NIT prote
gb_pr:BC020620	+ 479.92	679.92	1.6e-29	965	BC020620 Homo sapiens, NIT prot
gb_ba:AF005512	+ 478.50	668.82	6.9e-29	3180	AF005512 Nostoc commune tRNA-L
gb_ba:AP003587	+ 478.50	668.82	1.3e-26	34350	AP003587 Nostoc commune tRNA-L
gb_ro:BC020153	+ 474.50	672.62	4.2e-25	1063	BC020153 Mus musculus, NIT prot
gb_ro:AF284573	+ 474.50	670.92	5.2e-25	1292	AF284573 Mus musculus, NIT prot
gb_ba:AL646071	+ 462.00	608.95	1.5e-25	203050	AL646071 Ralstonia solanaceae
gb_sts:G72919	+ 465.00	646.62	1.2e-27	847	G72919 MARC 2849-2850:991933517

gb_ba:SMES91791	- 444.50	579.32	6.6e-24	340900	AL591791 Sinorhizobium me
gb_ba:AJ414158	+ 441.50	578.26	7.6e-24	235050	AJ414158 Yersinia pestis
gb_ba:AE004129	+ 431.50	590.66	1.5e-24	10881	AE004129 Vibrio cholerae C
gb_ba:AE005759	+ 427.50	583.54	3.9e-24	12748	AE005759 Caulobacter cresc
gb_pl:AF284572	+ 419.00	589.87	1.7e-24	1520	AF284572 Saccharomyces cere

seq\_name: gb\_ro:AF069988

seq\_documentation\_block:

LOCUS AF069988 1338 bp mRNA linear ROD 23-JUL-1998

DEFINITION Mus musculus nitrilase 1 (Ntl1) mRNA, complete cds.

ACCESSION AF069988

VERSION AF069988.1 GI:3228667

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM house mouse.

REFERENCE 1 (bases 1 to 1338)

AUTHORS Pektarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,

Tillib,S., Draganescu,A., Wermuth,P., Rotman,J.H., Huebner,K.,

Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

Nitrilase and fhit homologs are encoded as fusion proteins in

Drosophila melanogaster and Caenorhabditis elegans

Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

9837986

REFERENCE 2 (bases 1 to 1338)

AUTHORS Pektarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,

Tillib,S., Draganescu,A., Wermuth,P., Rotman,J., Huebner,K.,

Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

Direct Submission

Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson

Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

Location/Qualifiers

1. 1338

/organism="Mus musculus"

/db\_xref="taxon:10090"

/chromosome="1"

/map="1q21-q23"

1. 1338

/gene="Ntl1"

58. 1029

/gene="Ntl1"

/product="nitrilase 1"

/protein\_id="AAC40185.1"

/db\_xref="GI:3228668"

/translation="MGFTIRPPHQLCTGRLRLRPVLTQPRPRMSSSTSMELPL

VAVCOVTSFNRKQENKTCALVEQEARAGACAPLEAFDFIARNDAETLLISEPLN

GDLGGYSGIARREGGIMTISGSHRGGDEMNOKIYNCHVILNLSGVSVAARNTL

CDVETGGGMRMSNTTKRGTELEPPVKTACVAGIATCDMRFPELSKLAAAGAEI

LTVPSAFGSGTGAHWEVLILRAKATISQCVIAAOCGRHETRASTGHSKVVDPWGT

VVARGSGPGLCARIDLHFLQMRHLPVFOHRRDPLVYSLGSLPLS"

BASE COUNT 347 a 335 c 330 g 326 t

ORIGIN

alignment\_scores:

Quality: 1731.00 Length: 323

Ratio: 5.376 Gaps: 0

Percent Similarity: 99.690 Percent Identity: 99.690

alignment\_block:

US-09-357-675C-22 x AF069988 ..

Align seg 1/1 to: AF069988 from: 1 to: 1338

1 MetLeuGlyPheIleThrArgProProHnSgInLeuLeuCysThrGlyTyr 17

|||||

58 ATGCGGGGCTTATACACAGGCTCTCTCACACCTGTGTGACGGGATA 107

|||||

17 TArgLeuTauArgIleProValLeuCysThrGlnProArgProArgThrM 34

|||||

```

108 CCGATTACTTCGAACCCAGTACTTTGTACTACGCCAGCCAGACCA 157
34 etSerSerSerThrSerTrpGluLeuProLeuValAlaValCysGluVal 50
158 TGTCTCATCACTTCTCTGGAGAGCTGCCCTGGTGGCTGTGTCCAGGTA 207
51 ThrSerThrProAsnLysGlnGluAsnPhelSerThrcysAlaGluLeuVa 67
208 ACATCAACACCAACAGCAAGAGACACTTAAACATGTGCTGAGTTGGT 257
67 LGlnGluAlaAlaIarGluGlyAlaCysLeuAlaPheLeuProGluAlaP 84
258 TCAAGAGAGCTGCGACACTGGCTGTGCTTGTCTTGTCTGAGGAT 307
84 heaPheIleAlaIarGAsnProAlaGluThrLeuLeuLeuSerGluPro 100
308 TTGACTTTATTGACAGAACCCCTCCGAGACATTACTCTGTCCGACCA 357
101 LeuAsnGlyAspLeuLeuGlyGlnTrpSerGlnLeuAlaArgGluCysG 117
358 CTGAATGGGGATCTTTGGCCAAATATAGCCAGCTTGGCAGGAATGTGG 407
117 YLIETrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGlu 134
403 AATCTGGCTGTCTTGGGCGGTTTCCAGACCGTGGCCAAAGACTGGAGC 457
134 LnaAsnGlnLysIleTrpAsnCysHisValLeuLeuAsnSerLysGlySer 150
453 AGAATTCAGAAATCTACAAATGTCAATGCTTTGAAACAGCAAGGAGTCA 507
151 ValValAlaSerTrpArgLysThrHisLeuCysAspValGluIleProG 167
508 GTATGGCGCAGTTACAGGAAGACACATCTGTGCATGTAGAGATCCAGG 557
167 YGlnGlyProMetArgGluSerAsnTrpThrLysProGlyGlyThrLeu 184
558 TCAGAGGCGCGATGTAGAGAAACCACTATACCAACCTGGAGGACTCTTG 607
184 LrpProValLysThrProAlaGlyLysValGlyLeuAlaIleCysTrp 200
608 AGCCACCTGTCAAGACACCGGCTGGCAAGGTGGTGTACGACATCTGTAT 657
201 AspMetArgPheProGluLeuSerLeuLysLeuAlaGlnAlaGlyAlaG 217
658 GACATGGGGTCCCTGAACCTTTTGAATTTGGCTCAAGCTGGGGCAGA 707
217 uLIELeuThrTrpProSerAlaPheGlySerValThrGlyProAlaHis 234
708 AATCTTACTTATTCCTTCAGCCTTTGGATCTGTACAGGTCCAGCCACT 757
234 rPgIuValLeuLeuAlaArgAlaIleGluSerGlnCysTrpValIle 250
758 GGGAGGTGCTGCTGGGCGCGCCGACATGATCTCACTGCTATGATATA 807
251 ALaAlaIaGlnCysGlyArgHisHisGluThrArgAlaSerTrpGlyHis 267
808 GCACAGCGAGCGTGTGAGCGCCACCATGAACAGAGAAATTAATGAGCA 857
267 sSerMetValValAspProTrpGlyThrValValAlaArgCysSerGlu 284
858 TAGCATGTGTGTTACCGCTGGGCGACAGTGTGGCCGCGCTCCGAG 907
284 LrpProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuGlnGlnMet 300
908 GACCAAGGCTCTGCTTGGCTCGAAATGATCTCCACTTTACAAACAGATG 957
301 ArgGlnHisLeuProValPheGlnHisArgArgProAspLeuTrpGlySe 317
958 CGCCAAACACGCGCTGTGTTCAGACCGCAGACCTGACTATGGAG 1007
317 rLeuGlyHisProLeuSer 323
1008 TCTGGGTATCCACTCTCT 1026

```

seq\_name: gb\_01:BC021634

seq\_documentation block:

LOCUS BC021634 1365 bp mRNA linear ROD 22-JAN-2002

DEFINITION Mus musculus, nitrlase 1, clone MGC:13825 IMAGE:4008543, mRNA,

complete cds.

ACCESSION BC021634

VERSION BC021634.1 GI:18204912

KEYWORDS MGC.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Authors Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1365)

Strausberg, R.

Direct Submission

Submitted (14-JAN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

Contact: amadan@systemsbiology.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRK Plate: 18 Row: e Column: 20

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 6754855.

Location/Qualifiers

1..1365

/organism="Mus musculus"

/db\_xref="LocusID:27045"

/db\_xref="GI:18204913"

/clone="MGC:13825 IMAGE:4008543"

/tissue.type="Mammary tumor metastasized to lung."

/note="Vector: PCMV-SPORT6"

59..1030

/codon\_start=1

/product="nitrlase 1"

/protein\_id="AAH21634.1"

/db\_xref="GI:18204913"

/translation="MIGETTRPPOHLCITGYRLRTPVLCTQPPRTSSSYSEWELP

VAVCVSTSPNKOENKTCALVQEARAGACLAFLPEAFDITANPAATLLSPRL

GLDLOYSQLARECGIWLISGPHERGOMRONKTYNHVILNKGSVAYSRKTHL

CDVEITPGGPRAESNTKPGGTLPEPRVTPKAKVGLACIDYDRFPBELSLQAQAEI

LTYSNAFGSVTGAHMEVILARRALIESQCYVIAAQCGHHEHETRASYGSHVYDPMGT

VVAKSEGGVLCIALIDILHFDQMRHLPVFQHRRDLYGSLGHLPS"

BASE COUNT 368 a 334 c 331 g 332 t

ORIGIN

alignment\_scores:

Quality: 1723.00

Ratio: 5.368

Percent Similarity: 99.381

Percent Identity: 99.381

alignment\_block:

US-09-357-675C-22 x BC021634

Align seg 1/1 to: BC021634 from: 1 to: 1365

```

1 MetLeuGlyPheIleThrArgProHisGlnLeuLeuGlyThrGly 17
  |||
59 ATGCTGGGGCTTCATACCAAGGCTCTCCACCACTCCGTGTACCGGATA 108
17 ArgLeuLeuArgGleProValLeuGlyThrGlnProArgProArgThr 34
  |||
109 CCGATTACTTCGAACCCAGACTTGTACTACACCGCCAGCCAGAACCA 158
34 eSerSerSerThrSerTProGluLeuProLeuValAlaValCysGlnVal 50
  |||
159 TGTCTCATCAACTTCCTGGGAGCTGCCCTGGGGCTGTGTCCAGGTA 208
51 ThrSerThrProAsnLysGlnGlnAsnPheLysThrCysAlaGluLeuVal 67
  |||
209 ACATCAACACCAACAGACAGACACTTAAACATGTGTGAGTGTGT 258
67 LglnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaP 84
  |||
259 TCAAGAGGCTGCCAGACTGGGTGCTTGCTTGCTTGCTTGCTTGAGCAT 308
84 heaSpheIleAlaArgAsnProAlaGluThrLeuLeuLeuSerGluPro 100
  |||
309 TTGACTTTATTGACAGAAACCTGCCAGACATTACTCTCCGAACCA 358
101 LeuAsnGlyAspLeuLeuGlyGlnThrSerGlnLeuAlaArgGluCysG 117
  |||
359 CTGATGGGAGTCTTTGGGCCAATATAGCCAGCTTGCACAGGAATGGG 408
117 YLeuThrLeuSerLeuGlyGlnPheHisGluArgGlyGlnAspTrpGlu 134
  |||
409 AATGTGGCTCTCCTTGCGGGCTTCCAGACGCGGCCCAAGACTGGAGC 458
134 LAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlySer 150
  |||
459 AGAATCAGAAATCTACAATTGTGATGTGCTTGAACAGACAGAGATCA 508
151 ValValAlaSerTyrArgLysThrHisLeuCysAspValGluIleProG 167
  |||
509 GTATGGCCAGTTACAGAAACACATCTGTGCATGTAGAGATCCACGG 558
167 YGlnGlyProMetArgGluSerAsnTyrThrLysProGlyGlyThrLeu 184
  |||
559 TCAAGGGCCGATGAGAAACCACTATACCAACCTCGAGAGCACTCTG 608
184 LProProValLysThrProAlaGlyLysValGlyLeuAlaIleCysTyr 200
  |||
609 AGCCACCTGTCAAGACACCGGCTGCAAGGTTGGTCTAGCAATCTGTAT 658
201 AspMetArgPheProGluLeuSerLeuLysLeuAlaGlnAlaGlyAlaG 217
  |||
659 GACTGTGGGCTTCCTGAACTTCTTGAATTTGGCTCAAGCTGGGGCAGA 708
217 uIleLeuThrTyrProSerAlaPheGlySerValThrGlyProAlaHisT 234
  |||
709 AATPCTTACTTATCTTCAGCCTTGGATCTGTTCACAGGTCACGCCACT 758
234 TrpGluValLeuLeuArgAlaArgAlaIleGluSerGlnCysTyrValI 250
  |||
759 GGGAGGTGCTGCTGGGGCCCGCCCATTTGAATCTCAAGTCTAATGATA 808
251 AlaAlaAlaGlnCysGlyArgHisGlnIleThrArgAlaSerTyrGlyH 267
  |||
809 GCACACACCGAGTGTGAGACCGCACCATGMAACAAGACAGCAATTTATG 858
267 sSerMetValValAspProTrpGlyThrValValAlaArgCysSerGlu 284
  |||
859 TAGCATGTGTTGACCCGTTGGGACACAGTGTGTGGCCGCTGCTCCAGG 908
284 LProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuGlnGlnMet 300
  |||
909 GACCAGGCTCTGCTGCTGCAATGTATCTCCACTTCTACACAGATG 958

```

```

301 ArgGlnHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 317
  |||
959 CGCCACACCTGCTGCTGTTCACAGACCGCAGACCTGACTTATGGCAG 1008
317 rLeuGlyHisProLeuSer 323
  |||
1009 TCTGGGTATCCACTCTCT 1027
seq_name: gb-pr:AF069987
seq_documentation_block:
LOCUS AF069987 1385 bp mRNA linear PRI 23-JUL-1998
DEFINITION Homo sapiens nitrlase 1 (NITL) mRNA, complete cds.
ACCESSION AF069987
VERSION AF069987.1 GI:3228665
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1385)
Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrlase and fhit homologs are encoded as fusion proteins in
drosophila melanogaster and caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
98337986
MEDLINE
2 (bases 1 to 1385)
Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES
source
1..1385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q21-q22"
1..1385
/gene="NITL"
77..1060
/gene="NITL"
/feature="NITL"
/codon_start=1
/product="nitrlase 1"
/protein_id="AAC39907.1"
/db_xref="GI:3228666"
/translation="MUGFTTRPFRRLSLCPGLRIPQLSVCACPRPRAMAISSSC
ELPLVAVCOVSTPRDKQONFKTCAELVREARAGACIAPLEAFDEIARDPARTLHS
EPLGKLEEVTLAREGGLMTSLGFGHERGQWETQKLYNCHVLNLSGAVATYR
KTHLCVLEIPFGGPMCESTMPGSPLESFVSYPARKIGLAVCYDHRPFLSLAQA
GAEILLYPSAIFGSIITGPANHEVILRRALITQCYVVAACCGSHHRRKRSYSGHSMYD
PMGTIVARSGEGCLCLARDLNTYLQRLRHRLVFQHRPDLTGNLGHPLS"
BASE COUNT 322 a 382 c 357 g 324 t
ORIGIN
alignment_scores:
Quality: 1487.00 Length: 327
Ratio: 4.828 Gaps: 2
Percent Similarity: 94.190 Percent Identity: 84.404
alignment_block:
US-09-357-675C-22 x AF069987 ..
Align seg 1/1 to: AF069987 from: 1 to: 1385
1 MetLeuGlyPheIleThrArgProHisGln.....LeuLeu 14
  |||

```

```

77 ATGCTGGGCTTCATCACCAGGCTCTCCACAGATTCCTGCTTCCTG 126
14 sThrGlyTyrArgLeuLeuArgIleProValLeuCysThrGlnProArgp 31
   ||| |||::: |||::: |||::: |||::: |||::: |||::: |||:::
127 TCCGGAGCTCCGATACCTCACTCACTCACTCACTCACTCACTCACTCA 176
31 roArghrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 46
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
177 CAGAGGCAATGAGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 226
47 ValCysGlnValThrSerThrProAsnLysGlnLysAsnPhelysthrCy 63
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
227 GTGTGCAGGATTAACATCAGCCAGACACACAGCACTTTAAACATG 276
63 sAlaGluLeuValGlnGlnLysAlaArgLeuGlnValAlaCysLeuAlaPheL 80
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
277 TGTGAGAGCTGTTGAGAGAGCTGCCACACTGGGTGCTGCTGCTGCTTCC 326
80 euProGlnAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 96
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
327 TGCCTGAGGCAATTTGACTTCATTCACAGGGAGCCCTGCAGAGAGCTACAC 376
57 LeuSerLurProLeuAsnGlyAspLeuLeuGlyLysIntySerGlnLeuAl 113
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
377 CTGTCTGAACCACTGGGTGGGAACTTTTGGAAATACACCCAGCTTGC 426
113 aArgGluCysGlyIleTrpLeuSerLeuGlyLysPheHisGlnArgGlyG 130
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
427 CAGGAGATGAGGAGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 476
130 LnsApTrpGlnLysAsnGlnLysIleTyrAsnCySHisValLeuLeuAsn 146
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
477 AAGACTGGGAGAGACTCAGAAATCTACAATGTCTCAGCTGCTCTGAAC 526
147 SerLysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspVa 163
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
527 AGCAAAAGGAGAGTAGTGCCCTTACAGAGAGACACATCTGTGTGAGCT 576
163 LgluIleProGlyGlnGlyLysPheMetArgGluSerAsnTyrThrLysPro 180
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
577 AGAGATTCACAGGAGGAGGCTATGTGTGAAGACACTCAACCTAGCTG 626
160 LysGlyThrLeuGluProProValLysThrProAlaGlyLysValGlyLeu 196
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
627 GGCACAGCTTGTGAGTCACTGTCAGCACACAGAGAGGAGATGTGCTGA 676
197 AlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysAlaGln 213
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
677 GCTGTGCTATGATGACATGCGGTCCCTGAACCTCTCTGCGCATTTGCTCA 726
213 nAlaGlyAlaGlnIleLeuThrTyrProSerAlaPheGlySerValThrG 230
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
727 AGCTGGAGCAGAGATTAATCTTACTTCTCAGCTTTTGATCCATTACAG 776
230 LyrProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluSerGln 246
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
777 GCCACAGCCCACTGGAGAGTGTGCTGCGGCGCCCTGCTATGAAACCCAG 826
247 CysTyrTrpAlaIleAlaAlaGlnCysGlyArgHisHisGlnThrArgAl 263
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
827 TGCATATATATGAGCAGCAGCAGTGTGAGCGCCACCATGAAGAAGAGAGC 876
263 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAla 280
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
877 AAGTTATGGCAGACAGCATGTGTGTAGACCCCTGGGGAACAAGTGTGCCCC 926
280 rGcYsSerGlnGlyLysProGlyLeuCysLeuAlaArgIleAspLeuHisPhe 296
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
927 GCTGCTCTGAGGGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 976
297 LeuGlnGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAs 313
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
977 CTGCGACAGTTGGCGCGACACTGCTGTGTCTCCAGCAGCGCAGGCTCTGA 1026

```

```

313 pleuTyrGlySerLeuGlnHisProLeuSer 323
|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1027 CCTTATGCGCAATGTGGTCAACCACACTGTCT 1057

seq_name: gb_ro:AF069985

seq_documentation_block:
locus AF069985 4481 bp DNA linear ROD 23-JUL-1998
DEFINITION Mus musculus nitrlase homolog 1 (Nt1l) gene, alternatively spliced
product, complete cds.
ACCESSION AF069985
VERSION AF069985.1 GI:3242979
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 4481)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrlase and fhl1 homologs are encoded as fusion proteins in
Drosophila melanogaster and Caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
2 (bases 1 to 4481)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

FEATURES
source
1..4481
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/map="1q21-q23"
join(541..606,1177..1263,1638..1889,2015..2118,2362..2495,
2626..2751,3392..3658)
/gene="Nt1l"
/note="alternatively spliced"
/product="nitrlase homolog 1"
541..3944
/gene="Nt1l"
541..606
/note="Nt1l"
/gene="Nt1l"
/note="alternatively spliced"
/number=1
join(605..606,1177..1263,1638..1889,2015..2118,2362..2495,
2626..2751,3392..3658)
/gene="Nt1l"
/note="alternatively spliced"
/codon_start=1
/product="nitrlase homolog 1"
/protein_id="AAC40184.1"
/db_xref="GI:3242980"
/translation="MIGRTPRPHDLCSTGYRLRPVLTQPRPRMSSSWELPL
VAVCOVSTPNKQENFTCAELVQDAKIGASCLAPFAFDIARNPETLILSEPLN
GDLGQYSOLNAEGCWLISLGFHERGDWDQONLYNCVHLNLSKGSVAVSYRTHL
CDVEILPGCGPMRESNVTKPGGTLPEPVTPAGKVGILACYLMRPPELSLKAQACAEI
LTPSAFGSVTPAHMEVLLRARIIESQCYUAAACGRNHETRASYSKSHMVDPMGT
VVARSEGRGLAIDIDHPLODMQNHLPVQHRHRDLYSLGHPLS"
863..986
/gene="Nt1l"
/note="alternatively spliced"
/number=2
1177..1263
/gene="Nt1l"
/number=3
1638..1889
/gene="Nt1l"
exon
exon
exon

```

```
/number=4
2015..2118
/genes="Nt1L"
exon
/number=5
2362..2495
/genes="Nt1L"
exon
/number=6
2626..2751
/genes="Nt1L"
exon
/number=7
3392..3944
/genes="Nt1L"
/number=8
BASE COUNT 1101 a 1132 c 1133 g 1115 t
ORIGIN

alignment_scores:
Quality: 1347.00 Length: 828
Ratio: 4.196 Gaps: 9
Percent Similarity: 38.768 Percent Identity: 38.406

alignment_block:
US-09-357-675C-22 x AF069985 ..

Align seg 1/1 to: AF069985 from: 1 to: 4481

2 LeuGlyPheLeuThrArgProProHisLeuLeuCysThrGlyTyrAr 18
|||||
1178 CTGGGCTTCATCACCAGGCTCTCCACCACTCCTGTGACCGGATACCG 1227
18 gLeuLeuArgLeuProValLeuCysThrGlnPro..... 29
|||||
1228 ATTACTTGAAATCCAGTACTTGTACTGACGCCAGTAACATATTTCGT 1277
29 ..... 29

1278 AATGTTCTCTCTGCGCATAGCTGCTTAGTTCGTTATTAACTTCA 1327
29 ..... 29

1328 TAGTGGAGACAAGATATCTCCAAACATGTAATGAGTCACTCAGTAG 1377
29 ..... 29

1378 CTCTAGGGTTATTATGCTCTGTGAAGATGAAACAAGTGTGGGGTCA 1427
29 ..... 29

1428 CTGTAATGCCAGACTCGAGAAAAAGAGGATCAACCAAGAGTTTAAGT 1477
29 ..... 29

1478 CATCCCAAGTTACATAGAAAGGCTGTCTGGGCTACAAAAGCCGTCT 1527
29 ..... 29

1528 CAAAAAGAAAAAGTCAAAAAGAAATAAAGTTCTGTGCTCCACCCCCAC 1577
29 ..... 29

1578 TTGACCTCTGAGTTCTCTAGATAGAAAATCTGAGGATGCTGCTGAG 1627
30 .....Arg..ProArgThrMetSerSerSerThrSerTrpGluLeuPro 43
|||||
1628 TCATTACAGGCCCAAGAACCATGCTCTCATCACTCTCTGGGAGTGCC 1677
44 LeuValAlaValAlaCysGlnValThrSerThrProAsnLysGlnLysAsn 60
|||||
1678 CTGGTGGCTGTGTGCCAGGTACATCAACCAACCAAGCAAGAACTT 1727
60 eLysThrCysAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysL 77
|||||
```

```
1728 TAAACATGTGCTGAGTTGGTTCAAGAGGCTCCAGACTGSGGTCTTCC 1777
77 euAlaPheLeuProGluAlaPheAspPheIleAlaArgAsnProAlaGlu 93
|||||
1778 TGGCCTTCTGCTGAGGCAATTTGACTTTATTCAGCAAGAACCTGCGGAG 1827
94 ThrLeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyGlnTyrSe 110
|||||
1828 ACATTACTCTCTCCGAAACCACTGAATGGGATCTTTTGGCCAAATATAG 1877
110 rGlnLeuAlaArg..Glu..... 115
|||||
1878 CCAGCTGCCAGGTACAAAGGTAGAGGTTGGGGGGGGGGGTTGTTGTTGT 1927
115 ..... 115

1928 GTAGAACTCCCTGTGTGAGCAGTGTCTGTGAATTTGCCACAGAGGGGTAG 1977
116 .....CysGlyIle 118
|||||
1978 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2027
119 TrpLeuSerLeuGlyIlePheHisGluArgGlyGlnAspTrpGluLys 135
|||||
2028 TGGTGTCTCTTGGGGGTTTCCACGAGCGTGGCCAAAGTGGAGCAGAA 2077
135 nGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLys..... 148
|||||
2078 TCAGAAATCTCAATTTGATGTGCTTTTGAACACAGAGGTGAGACGT 2127
148 ..... 148

2128 CATAGCCCTTTAGTTGCTCTTCCAGGCTTCTACTTCTTCTCACT 2177
148 ..... 148

2178 CAGATCTCAGCATTTCTTCAACTGCTTGGGCTTGGGGTTTATGCT 2227
148 ..... 148

2228 TTTGTTTCAGCCTACAAAGTTACTTTTGGAGAAAGCTAAGCTGAGACT 2277
148 ..... 148

2278 TGTAGAGTCTGTACTGATATTTCTTCTGTCTCTCTCTCTCTCTCTCT 2327
149 .....GlySerValAlaLase 154
|||||
2328 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2377
154 rTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPro 171
|||||
2378 TTACAGGAAGACACATCTGCGATGTAGAGATCCAGGTCAGGGCCAA 2427
171 eTArgGluSerAsnTyrThrLysProGlyGlyThrLeuGluProProVal 187
|||||
2428 TGAGAGAAAGCAACTATACCAAGCTGAGGCACTCTTGAAGCCACTGTCT 2477
188 LysThrProAlaGly..Lys..... 193
|||||
2478 AAGACACCGGCTGGCAAGTGGAGTTTGAAGAAGAAATGGGACTATGTT 2527
193 ..... 193

2528 GAAAGTGTACATCCCATTTGTGACGGCTAATGAAGGCACTCAGTG 2577
194 .....Ys 194
|||||
2578 TTCTGTTGAATGGGACAGGAAAGATGAGTGGGTGTTCTCATTTAGGT 2627
194 lGlyLeuAlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysL 211
|||||
2628 TGGTCTAGCAATCTGTTATGACATGCGGTTCCCTGAACCTTCTTTAAAT 2677
```

```

211 euAlaGlnAlaGlyAlaGlnIleLeuThrTyrProSerAlaPheGlySer 227
|||||
2678 TGCGTCAGAGCGGGGAGAGAAATACTTACTTATCCCTTCAGCCCTTGATCT 2727
228 ValThrGlyProAlaHisTyr..... 234
|||||
2728 GTTACAGGTCAGCCCACTGGAGAGTAAGAAATCCCTTCATTAACATA 2777
234 ..... 234
2776 TAGGCCTCCCTTAACTTACTTTAGTAGTGAGGAGACAGTGTACAAAC 2827
234 ..... 234
2828 CTAATATCCAGACACTTACGTGATAGCCACAGGTTCAAAAGCAGCATGCC 2877
234 ..... 234
2878 TTGACTAAAAAAGAAACAAATAGAACTGGGCACTGGGGTGACACAC 2927
234 ..... 234
2928 TTTAATCCACACTCAGAGAGCAGACAGAGTGATCTGTGATTGAG 2977
234 ..... 234
2978 GCCAGCCTGCTCTACAGAGTGGGTTCCAAGACAGCGGCTACACAGAG 3027
234 ..... 234
3028 AAACCTGTCTCAAAACAAAAACAAACCAACCACTTCAAAGAGTGGGG 3077
234 ..... 234
3078 CATGTGTGATGCTCTTTCATCTCAGACACTGTAGGCAAGGAGGCGCAT 3127
234 ..... 234
3128 ATCTCTGGGCTTAGGCCAGCAGGTTACACAGGAGACCTTATCTCA 3177
234 ..... 234
3178 AAAGTACAAAACACAAACACTCTCTCTTCTTCTGATTTCTTCCTTTT 3227
234 ..... 234
3228 CATCTGGTGAAGAAATTCATTCTTAGCTAGGTGAGTCATGTTAAGA 3277
234 ..... 234
3278 GTTAGTTATTAATCTAGTTGGTTAGTGAGAAAGATTTGGCCACAGTGC 3327
234 ..... 234
3328 CGCTTCTTGCACTTGCCATTAACGCTGATGTCTTGAGATCAACGCA 3377
235 .....Glu.ValLeuLeuArgAlaArgAlaIleGluSerGlnCys 247
3378 CGTCTCACTCTCAGGTGCTGCTGGCGGCCCGCCCATTTGAATCTCAGTGC 3427
248 TyrValIleAlaAlaAlaGlnCysGlyArgHisHisGlnThrArgAlaSer 264
|||||
3428 TATGTATATAGACAGCAGCAGTGTGAGCGCCACATGAAACAAGAGCAG 3477
264 rTyGlyHisSerMetValValAspProTyrPglyThrValValAlaArgC 281
|||||
3478 TTTAGGCCATATGATGTGTGTGACCCCTGGGACACAGTGTGGCCGCT 3527
281 yssSerGlnGlyProGlyLeuGlyLeuAlaArgIleAspLeuHisPheLeu 297
|||||
3528 GCTCGAGGAGGAGCGCTCTGCTGCTGCAATTAATCTCACTTCTCTA 3577

```

```

298 GlnGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAspLe 314
|||||
3578 CAACAGATGGCGGCAACACCTGCTGTCTCCAGCAGCCGACACTGACCT 3627
314 uTyGlySerLeuGlnHisProLeuSer 323
|||||
3628 CTATGGCAGTCTGGGTCACTCCTCTCT 3655
seq_name: gb_htg:AC084821
seq_documentation block:
LOCUS AC084821 211771 bp DNA linear HTG 29-JAN-2002
DEFINITION Mus musculus chromosome 1 clone rp23-395h6, WORKING DRAFT SEQUENCE,
1 ordered pieces.
ACCESSION AC084821
VERSION AC084821.24 GI:16390260
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211771)
REFERENCE 1 Deschamps, S., Gu, W. and Roe, B.A.
Mus musculus BAC Clone rp23-395h6
JOURNAL Unpublished
2 (bases 1 to 211771)
Deschamps, S., Gu, W. and Roe, B.A.
REFERENCE Direct Submission
AUTHORS Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
JOURNAL On Jan 29, 2002 this sequence version replaced gi:16076995.
COMMENT ----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 211771: contig of 211771 bp in length.
* Location/Qualifiers
FEATURES
source
1. 211771
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone_lib="rp23-395h6"
/clone_lib="RP23 Mouse BAC Library 23"
BASE COUNT 55622 a 48688 c 49207 g 56231 t 23 others
ORIGIN
alignment_scores:
Quality: 1343.00 Length: 824
Ratio: 4.197 Gaps: 9
Percent Similarity: 38.835 Percent Identity: 38.471
alignment block:
US-09-357-675C-22 x AC084821 ..
Align seg 1/1 to: AC084821 from: 1 to: 211771
2 LeuGlyPheIleThrArgProProHisGlnLeuLeuGlyThrGlyTyr 18
|||||
529 CTGGGCTTACACACAGGCTCTCACCACCACTCTGTGTACCGGATACCG 578
18 glauLeuArgIleProValLeuGlyThrGlnPro..... 29

```



```

|||||  |||||||  |||||||
579 ATACTTCGAACCCAGTACTTTGTACTCAGCCAGGTAACATATTTCGT 628
29 ..... 29
629 AATGTCCTCTCTGCGGCATAGCTGCTTAGTGTGTTATTACTTCA 678
29 ..... 29
679 TAGTGGAGACAAGATATCTCCAAACATGTAATGTAAGTCACTCAGTAG 728
29 ..... 29
729 CTCTAGGGTTATTATGCTCTGTAGAAATGAAACAAAGTCTGGGGTCC 778
29 ..... 29
779 CTGTAATGCCAGACATCGAAAAAAGAGATCAACAGAGATTAAAGT 828
29 ..... 29
829 CATCCCAAGTTACATAGAAAGCGCTGTCTGGCTACAAAGACCGTGTCT 878
29 ..... 29
879 CAAAAAGAAAAAGTCAAAAAAATRAAGTTTCTGCTCCACCCCCAC 928
29 ..... 29
929 TTGAGCCTTGAGTTTCTCTAGATAGAAATCTAGGATACCTGCTGTG 978
30 .....Arg..ProArgThrMetSerSerThrSerThrProAlaIleuPro 43
979 TCATTCAAGGCCGAGAACCATGTCATCACTTCCTGGAGCTGCC 1028
44 LeuValAlaValCysGlnValThrSerThrProAlaIleuValAsp 60
1029 CTGGTGGCTGTGTCCAGGTACATCAACACCAACAAAGCAAGAACTT 1078
60 eLysThrCysAlaGluLeuValGlnIleuAlaArgLeuGlyAlaCysL 77
1079 TAAACATGTCTGAGTGGTTCAAGAGCGTCCAGACTGGGTGCTGCC 1128
77 euAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAla 93
1129 TGGCCTTTCCTGAGGCAATTGACTTTATTGACGAAACCTGCCGAG 1178
94 ThrLeuLeuSerGluProLeuAsnGlyAspLeuGlyIleuValIle 110
1179 ACATTACTCTCTGCCAACCACCTGAATGGGATCTTTGGCCAAATATAG 1228
110 rGlnLeuAla..... 113
1229 CCAGCTTGCCAGGTACAAAGGTAGGGTTTGGGGGGGGGGTGTGTGT 1278
113 ..... 113
1279 GTAGACTCCCTGTTGAGCAGTGTCTGATTGCCACAGAGGGGGTGA 1328
114 .....ArgGluCysGlyIleuTyrPle 120
1329 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1378
120 uSerLeuGlyGlyPheHisGluArgGlyGlnAspTyrPleGluGlnAsn 137
1379 GTCTTGGGGCGTTTCCAGAGCGTGGCCAAAGACTGGGAGCAGAAATCA 1428
137 ySIIeTyrAsnCysHisValLeuLeuAsnSerLysGly..... 149
1429 AAATCTACAAATGTGATGTGCTTTGAAACAGCAAGGGTGAGACGTCA 1478
149 ..... 149
1479 CCCTTAGTTGCTCTTCCAGGCTTCTCTACTCTTCTCTCAACTCAGATC 1528
149 ..... 149
1529 TTCAGCATTTCTTTCTCAACTGCTTGGCTTGGGTTTATGCTTCTGT 1578
149 ..... 149
1579 CTCAGCTTCAAGTTACTCTTTGGGAGAGAGCTAAGCTGACGACTTGA 1628
149 ..... 149
1629 GTGTCTGACTAGATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1678
150 .....Ser..ValValAlaSerTyrArgLysTh 158
1679 CTCTCTCTCTCTCTCTCTCTCTAGATCAGTATGGCCAGTACAGAGAAC 1728
158 rHisLeuGlyAspValGluIleProGlyGlnGlyProMetArgGlySer 175
1729 ACATCTGTGCATGTAGAGATCCAGGTGAGGGGGCGATGAGAGAAAGCA 1778
175 snTyrThrLysProGlyGlyThrLeuGluProProValLysThrProAl 191
1779 ACTATACCAAGCTGAGGCACTCTTGAGCCACCTGTCAAGACACGGCT 1828
192 Gly..Lys..... 193
1829 GGCAGAGTGGAGTTTGGAAAGAAATGGACTATGTGAAAAAGTGTAA 1878
193 ..... 193
1879 CACCCCATTTGGTCAGGCTAATGAAAGCACTCAGTCTCTGTGAATG 1928
194 .....ValGlyLeuAlaI 198
1929 GGACAGGAAGATGAGTGGTTTCTCATTTTCAGTTGGCTTGCAT 1978
198 eCysTyrAspMetArgPheProGluLeuSerLeuAlaGlnAla 215
1979 CTGTATGACATGCGGTCCCTGGAACCTTCTTGAATGGCTCAAGCTG 2028
215 yAlaGluIleLeuThrTyrProSerAlaPheGlySerValThrGlyPro 231
2029 GGGCAGAAATACTTATCTTCATCCCTTGAGCTTGGATGATACAGGTCA 2078
232 AlaHisTrp..... 234
2079 GCCCATGAGGATAGAGAAATGCTCTTTCATATCATATAGGCTTCCTT 2128
234 ..... 234
2129 AACCTTACTTTTAGTGTGGAGCAGAGTGTACAAACCTAATCCAAAGC 2178
234 ..... 234
2179 ACTTCAGTGATAGCCACAGTTCAAGCCAGCATGCTTGACTATAAAAA 2228
234 ..... 234
2229 GAAACAAATTAAGAACTGGGAGTGGGGGTGCACACCTTTAATCCAGC 2278
234 ..... 234
2279 ACTCAGAGAGCAGAGCAGGTGATCTGAGTTTGAGGCCAGCGTGTGTC 2328
234 ..... 234
2329 TACAGATGGGTTCCAGAGACAGCGCTACACAGAAACCTGTCTC 2378
234 ..... 234
2379 AAAAAAAGAAAGCAACCAACACTTCAAAAGTGGGCTGTGTGTGCA 2428

```

```

234 ..... 234
2429 TGCTCTCATCTCAGCACTGTGAGGCAAGGACAGATCTGTGGGT 2478
234 ..... 234
2475 CTAGACGACGAGGGTACACAGGAGACCTTATCTCAAAAATACAAAC 2528
234 ..... 234
2525 ACAACACTCCTCTTCTACTGATTTTCTCTTCTTCTCATCTGTGAGG 2578
234 ..... 234
2579 AAATTCATTCTTAGCTAGGTGATGATCATGTTAAGAGTTAGTTATTA 2628
234 ..... 234
2629 CTCTAGTTGGTTAGTGAGAGATTGGCCACAGTCCCGCTTCTGTCA 2678
235 ..... G1 235
2679 CTTCCTATTAAACGCTGTATGTTCTGAGTCAACGACGCTCATCTC 2728
235 u.VallleuleuAryAlaAryAlaIleGluSerGlnCysTyrValIleAla 251
2729 AGGTGCTGTGCGGGCCGCCCATTTGATCTGATGCTGTAATGTAATGCA 2778
252 ALAAlaGlnCysGlyArgHisHisGluThrArgAlaSerTyrGlnHis 268
2779 GCACGCGAGGTGTGACGCCACCATGAACAAAGAGTTATGCGCATAG 2828
268 rmetValValAspProTrrpGlyThrValAlaArgCysSerGlnGlyP 285
2829 CATGCTGTGACCCGTTGGGCGACAGTGTGCCCCGCTGCCGAGGAC 2878
285 roGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuGlnGlnMetArg 301
2879 CAGGCTCTGCTGCTGCTGATGATCTCCACTTTCACAAACAGATGCGC 2928
302 GlnHisLeuProValPheGlnHisArgArgProAspLeuTyrGlySerLe 318
2929 CACACCTGCTGTGTTCACACCGCAGACTGACTTATGACAGTCT 2978
318 uGlyHisProLeuSer 323
2979 GGGTCATCCACTCTCT 2994

```

seq\_name: gb\_htg:AC087229

seq\_documentation\_block:

LOCUS AC087229 213203 bp DNA linear HTG 05-FEB-2002

DEFINITION Mus musculus chromosome 1 clone rp23-191a19, WORKING DRAFT  
SEQUENCE, 4 unordered pieces.

ACCESSION AC087229 GI:18497119

VERSION AC087229.15 HTG: HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS house mouse.

SOURCE Mus musculus.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 213203) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS Jiang X., Song L., Gu W. and Roe B.A.

JOURNAL Unpublished

2 (bases 1 to 213203) Jiang X., Song L., Gu W. and Roe B.A.

COMMENT Submitted (19-DEC-2000) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
On Feb 5, 2002 this sequence version replaced gl:18390237.

----- Genome Center  
Center: Department of Chemistry And Biochemistry  
The University of Oklahoma  
Center code:UOKNOR

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 11373: contig of 11373 bp in length  
\* 11374 11473: gap of unknown length  
\* 11474 43477: contig of 32004 bp in length  
\* 43478 43577: gap of unknown length  
\* 43578 126526: contig of 82949 bp in length  
\* 126527 213203: contig of 86577 bp in length.  
Location/Qualifiers

FEATURES  
source 1. 213203  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/clone="rp23-191a19"  
/clone\_lib="RPI Mouse BAC Library 23"

BASE COUNT 57858 a 50465 c 49150 g 55425 t 305 others  
ORIGIN

alignment\_scores:  
Quality: 1320.50 Length: 824  
Ratio: 4.127 Gaps: 10  
Percent Similarity: 38.835 Percent Identity: 38.350

alignment\_block:  
US-09-357-675C-22 x AC087229 ..

Align seg 1/1 to: AC087229 from: 1 to: 213203

```

2 LeuGlyPheIleThrArgProPheHisGlnLeuLeuCysThrGlyTyr 18
48196 CTGGGCTTCATCACACGAGCCCTCCACCAACTCCTGTGTACCGGATACCG 48245
18 gLeuLeuArgIleProValLeuCysThrGlnPro..... 29
48246 ATTACTTCGAACCCAGTACTTGTACTCAAGCCAGGTAACATATTTCGT 48295
29 ..... 29
48296 AATGTTCTCTCTCGGCGCATAGCTGCTTAGTTCGTTTATTAACTTCA 48345
29 ..... 29
48346 TAGTGGAGACAAGATATCTCCAAACATGTATAGTGAATCCAGTAG 48395
29 ..... 29
48396 CTCTAGGGTTATTATGCTCTGTGAAGATGAACAAACAACTTCTGGGGTCCAC 48445
29 ..... 29
48446 CTGTATGCGACGACTCGAAAAAAGAGAGATCAACGAGAGTTTAAGT 48495
29 ..... 29
48496 CATCCCGATTACATAGAAAGGCTGTGTGGCTACAAAAAGACGCTCT 48545
29 ..... 29
48546 CAAAAAGAAAAAGTCAAAAAAATAAAGTTTCTGGCTCCACCCCCAC 48595

```

```

29 ..... 29
48396 TTGAGCCTCTGAGTTTCTCTAGATAGAAATCTGAGATACTGCTGTG 48645
30 .....Arg..ProArgThrmSerSerThrmSerTrpGluLeuPro 43
48646 TCATTCAAGAGCCGAGAACCATGCTCATCACTCTCGGAGGTGCCC 48695
44 LeuValAlaValCysGlnValThrSerThrmProAsnLysGlnGlnAsp 60
48696 CTGGCGCTGTGTGCCAGTAACTCAACACCAACCAAGCAAGAACACTT 48745
60 elysThrcysAlaGluLeuValGlnGlnAlaAlaArgLeuGlyAlaCysL 77
48746 TAAACATGCTGAGTGTGTTCAAGAGCTGCCAGACTGGGTGCTTCC 48795
77 euAlaPheLeuProGluAlaPheAspPheAlaAlaArgAsnProAlaGlu 93
48796 TGGCCTTCTGCGTGAAGCACTTGTACTTATTGCAAGAACCTGCCAG 48845
94 ThrLeuLeuLeuSerGluProLeuAsnLysPheLeuGlyLysTrpSe 110
48846 ACATTACTCTGCTCCGAACCACTGATGGGATCTTTGGGCCAATATAG 48895
110 rGluLeuAlaArg..Glu..... 115
48896 CCAGCTTCCAGGTACAAAGGTAGGGTTGGGGGGGGGTTGTGTG 48945
115 ..... 115
48946 TAGAACTCCCTGTTGAGCAGTGTCTGTGATTGCCACAGAGGGGTAGAG 48995
116 .....CysGlyLLeuTrpLeu 120
48996 TTGTGTGTGTGTGTCTCCGCCGCCGCCGCCAGGGAATGTGAATCTGGCTG 49045
121 SerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnAsnGly 137
49046 TCTTGGGGGGGTTTCCAGAGCGGTGCCAAGACTGTGGAGCAAAATCGAA 49095
137 sIlLeuTrnCysHisValLeuLeuAsnSerLysGly..... 149
49096 AATCTACAATTTGTCTATGTCTTTGAACAGCAGAGGTGAGAGCTATAGC 49145
149 ..... 149
49146 CCTAGTTTGCCTTCCAGGCTTCTACTTCTCTCAACTCAGATCT 49195
149 ..... 149
49196 TCAGCATTTCTTCTCAACTGCTTGCCCTTGGGGTTTATGCTCTCTTC 49245
149 ..... 149
49246 TCAGCCTACAGTTACTCTTGGGAGAGAGCTAGAGCTGAGCTGTAGAG 49295
149 ..... 149
49296 TGTCTGTACTAGATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTC 49345
150 .....Ser..ValValAlaSerTrpArgLysThr 158
49346 TCTCTCTCTCTCTCTCTAGAGATGAGTGGCCAGTTACAGAGAGACA 49395
159 HisLeuCysAspValGluLeuProGlyGlnGlyPrometLarGluSerAs 175
49396 CATCTGTCGATGATGAGATCCAGCTCAGGCGCGATGAGAGAAAGCA 49445
175 nTyTrpThrLysProGlyGlyThrLeuGluProProValLysThrProAlaG 192
49446 CTATACCAAGGCTGGAGGACACTTGTAGGCACTGTCAAGACACCGGCTG 49495
192 Ty.Lys..... 193
49496 GCAAGGTGGAGATTGGAAAGAAATGGGACTATGTGAAGAAGTGTAC 49545
193 ..... 193
49546 ACCCCCATTTGGTGCAGGCTAATGAAGCACTAGTGTCTGTGAATGG 49595
194 .....ValGlyLeuAlaIle 198
49596 GACAGGGAAGATGAGTGGGTGTCTTCATTTCAAGTTGGTGTCTACCAATC 49645
199 CysTrpAspMetArgPheProGluLeuSerLeuLysLeuAlaGluAlaG 215
49646 TGTATGACATGCGGTTCCTGAACTTCTTTGAATTTGGCTCAAGCTGG 49695
215 yAlaGluLeuLeuThrTrpProSerAlaPheGlySerValThrGlyPro 232
49696 GGCAGAAATACTTACTTATCCTTCAGCCTTGTGATCTGTACAGGTCCAG 49745
232 IaHisTrp..... 234
49746 CCCACTGGAGGTAAAGAAATGCTCTTCATATATATAGGCTCTTTA 49795
234 ..... 234
49796 ACCTTACTTTTAGTGTGTGGAGCAGATGTTACAAACCTAATCCAGCA 49845
234 ..... 234
49846 CTTAGTGGATAGCCACAGGTTCAAAAGCAGCATGCTTGACTAAAAAG 49895
234 ..... 234
49896 AAACAAATATAGAACTGGGAGTGGGGTGACACCTTTAATCCAGCA 49945
234 ..... 234
49946 CTCAGAGAGCAGACAGAGTGTCTGTAGTTTGAAGCCACCTGTGTCT 49995
234 ..... 234
49996 ACAGATGGGTTCCAGAGACAGCCAGGGCTACACAGAGAAACCTGTCTCA 50045
234 ..... 234
50046 AAACAAAAAAGCAACCAACACCTTCAAGAGTGGGGCATGTGTGATCAT 50095
234 ..... 234
50096 GCTTTCATCTCAGACACTGTGAGCAAGAGCAGAGATCTGTGGGTTC 50145
234 ..... 234
50146 TAAGCCAGCCAGGTTACACAGAGGAGACTTATCTCAAAAGTAAACA 50195
234 ..... 234
50196 CAACACTCCCTCTTCTACTGATTTCTCTCTCTCTCTCTCTCTCTCT 50245
234 ..... 234
50246 AATTATTTCTTACGAGGTGAGTCATGTTAAAGATTAAGTTATTAAC 50295
234 ..... 234
50296 TCTAGTGTAGTGAAGATTGGGCCACAGTGCCGCTCTGTGTAC 50345
235 .....Glu 235
50346 TTGCCTATTAAACGCTGATGTCTGTGAGTCAACGACGCTCATCTCA 50395
236 'ValLeuLeuArgAlaArgAlaIleGluSerGlnCysTrpValIleAla 252

```

50396 GGTCTGCTGCGGGCCGCCCATTTGAATCTCAGTGTATGTAATGACAG 50445  
252 laalagincysglYArghishisluhrargalaserTyglyHisser 268  
|||||  
50446 CAGGCGAGTGTGGAGCCACCATGAACAGACAGATATGGCATATGC 50495  
269 MetValValAspProTTPGlyThrValValAlaArgCysSerGluGlyPr 285  
|||||  
50496 ATGTGGTGTGACCCGTGGGCGACAGTGTGGCCGCTGCTCCGAGGAGCC 50545  
285 OGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuGlnIleMetArg 302  
|||||  
50546 AGCGCTCTGCTTCTGCAATTGATCTCCACTTCTACACAGATGCGCC 50595  
302 InHisLeuProValPheGlnHisArgArgProAspLeuTyGlySerLeu 318  
|||||  
50596 AACACCTGCGCTGTTCACAGCCGACACTGACTTATGACAGTCTG 50645  
319 GlyHisProLeuSer 323  
|||||  
50646 GGTATCTCCTCTCT 50660  
seq\_name: gb\_pr:AF069984  
seq\_documentation\_block: 4079 bp DNA linear PRI 23-JUL-1998  
LOCUS AF069984  
DEFINITION Homo sapiens nitrlase homolog 1 (NIT1) gene, alternatively spliced  
product, complete cds.  
ACCESSION AF069984  
VERSION AF069984.1 GI:3242977  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 4079)  
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganesu,A., Wermuth,P., Rothman,J.H., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Nitrilase and fhl homologs are encoded as fusion proteins in  
drosophila melanogaster and caenorhabditis elegans  
Proc.Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
JOURNAL 9837986  
MEDLINE 2 (bases 1 to 4079)  
REFERENCE  
AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganesu,A., Wermuth,P., Rothman,J., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Direct Submission  
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tench St., Philadelphia, PA 19107, USA  
FEATURES  
source  
1..4079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q21-q22"  
join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,  
2609..2734,3001..3569)  
/gene="NIT1"  
/note="alternatively spliced"  
/product="nitrlase homolog 1"  
635..3569  
/gene="NIT1"  
635..680  
/gene="NIT1"  
/note="alternatively spliced"  
/number=1  
join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,  
2609..2734,3001..3267)  
/gene="NIT1"  
/note="NIT1: alternatively spliced"  
/codon\_start=1  
/product="nitrlase homolog 1"

/protein\_id="AAC39901.1"  
/db\_xref="GI:3242978"  
/translation="MLGFTTRPPHFRFLSLCPGLRIPQLSVLCAOPRRPRAMASSSC  
EPLVAVCOVSTSPKQONFKTCAELVREARADLAPLPAEPFIADPAETLHL  
EPLGKLEETVQLAREGLMISLGGFHEHGGDMROTKIYNCHVNLKGVAVTYR  
KTHLCEVETPCGGPACNESYTPGSPLESVPSTPAGKICLAVCYDMRPELSIALAQA  
GAEILTPSAFESITGPAAHEVLRARALETCYVAAACGRHHEKRASTYGHMAYD  
PAGTVVARSSEGPGLARIDILNLRRLRPVQHRRLDLYGNLGHPLS"  
681..711  
/gene="NIT1"  
/note="1A: alternatively spliced"  
712..741  
/gene="NIT1"  
/note="1B: alternatively spliced"  
940..1053  
/gene="NIT1"  
/note="1C: alternatively spliced"  
1288..1383  
/gene="NIT1"  
/number=2  
1636..1890  
/gene="NIT1"  
/number=3  
2011..2114  
/gene="NIT1"  
/number=4  
2321..2454  
/gene="NIT1"  
/number=5  
2609..2734  
/gene="NIT1"  
/number=6  
3001..3569  
/gene="NIT1"  
/number=7  
BASE COUNT 940 a 1082 c 1043 g 1012 t 2 others  
ORIGIN  
alignment\_scores:  
Quality: 1211 50 Length: 660  
Ratio: 3.946 Gaps: 10  
Percent Similarity: 46.515 Percent Identity: 41.667  
alignment\_block:  
US-09-357-675C-22 x AF069984 ..  
Align seg 1/1 to: AF069984 from: 1 to: 4079  
2 LeuGlyPheIleThrArgProProHisGln.....LeuLeuCysTrh 15  
|||||  
1289 CTGGGCTTCATCACCAGGCGCTCCACAGATTCTGTCCTGTGTGCC 1338  
15 rlyTyArgLeuLeuArgIleProValLeuCysThGlnPro..... 29  
|||||  
1339 TGGAGTCGCGATACCTCAACCTCAGTACTTGTGCTCAGCCAGTAAAC 1388  
29 ..... 29  
1389 ACGTTTGTGTGTCTCAGTGCCTGGCAGCTTACATGCTCAGTTGTTAA 1438  
29 ..... 29  
1439 ATGATAGTGGAGACACAGAGAGTGTCACTATCCACACATTGATTGGT 1488  
29 ..... 29  
1489 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAAGCATGATCAAAAG 1538  
29 ..... 29  
1539 GAATCCAGCTTTCCTGCTCTCCACTTGCACCCCTTAGCATTAATTTC 1588

```

30 .....Argpr 31
1589 CTTCCCTGCTGATGAATCTGAGAACTCCCTATGCTGTTTCACAGGCC 1638
31 oArgThrMetSer...SerSerThrSerTrpGluLeuProLeuValAlaIav 47
1639 CAGAGCCATGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1688
47 aLysGlnValThrSerThrProAsnLysGlnGluAsnPhelLysThrcys 63
1689 TGTGCCAGGTAACTCGACGCCAGCAAGCAACACTTAAACATGT 1738
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlnValCysLeuAlaPhle 80
1739 GCTAGCTGCTTCAGAGGCTGCCAGACTGGGTGCTGCTGCTGCTTCTCT 1788
80 uProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeuL 97
1789 GCCTGAGGCATTTGACTTCATTCACGGGACCTTCAGAGACGGTACACC 1838
97 euSerGluProLeuAsnGlyAspLeuGlnGlyIntYrSerGlnLeuAla 113
1839 TGTCTGAACCACTGGTGGGAACTTTTGAAGATACACCAGCTTCC 1888
113 ..... 113
1889 AGGTATCAGGAATAGCGAGGAGAGGTAGAATCTTTGTTGACAGTGT 1938
113 ..... 113
1939 CCTGGGTGCCAGATATGAGGATAGACCTTGAGAACTCAGTGAAGAT 1988
114 .....ArgLysGlyIleTrpLeuSerLeuGly 123
1989 TGTCAAGTCCCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2038
124 GlyPheHisGluArgGlyGlnAspTrpGluGlnAsnGlnLysIleTyrAs 140
2039 GGTTCCATGGCGGTGGCCAAAGCTGGGAGCAAGCTCAGAAATCTACAA 2088
140 nCysHisValLeuLeuAsnSerLysGly..... 149
2089 TTGTACAGCTGCTGCTGAACACCAAGGTAGACCTTATACCTTTACC 2138
149 ..... 149
2139 TGCCCTCTCCATGCTCTTCTACTAGATCTCCAGAAATGTTTCTCAAC 2188
149 ..... 149
2189 TCTATTCTTCTTGAACCAAGATTAGGGGTGCTCTACTTCACTTCTCTA 2238
149 ..... 149
2239 GCCATAAACAATCTCTCTCTCTGGGAGAGAGTAAGCAAGCTTTAGAAC 2288
150 .....Ser..ValValAlaSerT 155
2289 CCACACATGATATCTCTCTCTTCTTACTGTAGGCGCAGTAGGCGCCTT 2338
155 YrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 171
2339 ACAGGAAACACATCTGTGTGACGTAGAGATTCCAGGAGGAGGCGCTTAG 2388
172 ArgLysSerAsnTyrThrLysProGlyGlyThrLeuGluProProValL 188
2389 TGTGAAGCAACTCTACCATGCTGCGCCAGTCTTGAGTCACTGCTGACG 2438
188 sThrProAlaGlyLysValGlyLeu..... 196
2439 CACACACACAGGCAAGGTAGAGTTGTGAAGATGAGGAGGAGGAGACAG 2488
196 ..... 196

```

```

2489 GAATACCTTGAACCTGGCAGTAGAGATAGAAAGCCCTAAGAGAGGGGTA 2538
196 ..... 196
2539 ATGGAATATGATAGTATGATCTGTGACAAACAGAGCAGAGACTACTAAG 2588
197 .....AlaIleCysTyrAspMetArg 203
2589 TAGGCTGTTTTTCATTCACAGATTGGTCTAGCTGTCTGTATGACATGCGG 2638
204 PheProGluLeuSerLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuTh 220
2639 TTCCTGAACTCTCTCTGCACTGGCTCAAGCTGAGCAGAGATATCTTAC 2688
220 TyrProSerAlaPheGlySerValThrGlyProAlaHisTrpGluVal. 236
2689 CTATCTTCACGCTTTTGGATCATATACAGCCACCCACTGGAGGTAA 2738
237 Leu..... 237
2739 GATGATGCCCTTTTAAACATTAAGGGCTTTTCTTAACCTCATCTTCCC 2788
237 ..... 237
2789 CCTTGCCCTACAGTAAATTCCTTCCCTTTCCACCTAATGGGAAAC 2838
237 ..... 237
2839 TCATCCCCAGATATTCTCTCATGATAGTAAATAGTAATCATTC 2888
237 ..... 237
2889 TAGCAATTACCAAAATAGTACAAATGGGTAGATTGCTGTGAATGTCCC 2938
237 ..... 237
2939 TCACCTGTACCTTCCCACTATTGCTACTAGTACTTAAGTAAACACACA 2988
238 .....LeuArgAlaArgAlaIleGluSerGlnCysTyr 248
2989 TCTCATGCCACAGGTGTGTGGGGCCCGTGTATCGAAACCCAGTGTGA 3038
248 YrAlaIleAlaAlaGlnCysGlyArgHisHisGluThrArgAlaSerT 265
3039 TGTAGTGCACGACGACAGTGTGAGCGCCACCATGAGAAAGACGCAAGTT 3088
265 YrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 281
3089 ATGCCACACGATGTGTGAGACCCCTGGGGAACAGTGTGGCCGCTGTC 3138
282 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuG 298
3139 TCTGAGGGGCCAGGCTGTGCTTCCCGCAATPACCTCACTATCTGCG 3188
298 nGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAspLeuT 315
3189 ACAGTTGGCCGACACCTGCTGTGTCTCCAGCACGCGACGCGCTGACTCT 3238
315 YrGlySerLeuGlnHisProLeuSer 323
3239 ATGCCAATCTGGGTCAACCCACTGTCT 3264

```

seq\_name: gb\_hlg:AL354714\_4

seq\_documentation\_block:

WPCOMMENT

Sequence	split	into	7	fragments	LOCUS	AL354714	Accession	AL354714
Fragment	Name	Begin	1	End	11000			
AL354714_0		100001		210000				
AL354714_1		200001		310000				
AL354714_2		300001		410000				
AL354714_3		300001		410000				

AL354714\_4 400001 510000  
AL354714\_5 500001 610000  
AL354714\_6 600001 699820  
Continuation (5 of 7) of AL354714 from base 400001 (AL354714 Homo sapiens chromosome 1 c

## alignment\_scores:

Quality: 1211.50 Length: 660  
Ratio: 3.946 Gaps: 10  
Percent Similarity: 46.515 Percent Identity: 41.667

## alignment block:

US-09-357-675c-22 x AL354714\_4 ..

Align seg 1/1 to: AL354714\_4 from: 1 to: 110000

```
2 LeuGlyPheIleThrArgProProHisGln.....LeuLeuGlyTh 15
|||||
103858 CTGGGCTTCATCCAGGCGCTCTCACAGATTCTGTCCTGTGTGTC 103907
15 rGlyTyrArgLeuLeuArgIleProValLeuGlyThrGlnPro..... 29
|||
103908 TGGACTCCGGATACCTCACTCTCACTTGTGTCTCAGCCCAAGTAAC 103957
29 .....
103958 ACCTTTGTTGTCTCTCAGTGCCTGCGACCTAGATGCTAGTTGTTAA 104007
29 .....
104008 ATGGATATGTTGGAGACACAGAGTGTCACTATCCACACATTGATTGGT 104057
29 .....
104058 GAGCCCTACTAGCCCTGGTCAACGTGCCCTGTAGAGCATGATCAAAAG 104107
29 .....
104108 GAAGTCCAGCTTCTGCTCTCCACTTCACCTTACCCCTTAGATTAATTGG 104157
30 .....
104158 CTTCCTCTGCTCTATGAAATCTGAGAAATCTGCTATGCTGTACAGGCC 104207
31 oArgThrMetSer...SerSerThrSerTyrGluLeuProLeuValAla 47
|||||
104208 CAGAGCATGTCATCTCTCTCTCTCTGCGAACTGCCCTGTGTGCTG 104257
47 aLcysGlnValThrSerThrProAsnLysGlnIleAsnPhelysThrCys 63
|||||
104258 TGTGCCAGGTAAATCGACGCCAGACAAACGAACTTTAAATCATGT 104307
64 AlagIleuValGlnGlnIleAlaIleArgLeuGluValAlaCysLeuAlaPheLe 80
|||||
104308 GCTGAGGTGTTGTGAGAGGCTGCGACAGCTGGTCCGTGCTGCTTCC 104357
80 uProGluAlaPheAspPheIleAlaIleArgAsnProAlaGluThrLeuLeu 97
|||||
104358 GCCTGAGGCTTTGACTTCATTCGACGGGACCTGCGAGAGACGTTAAC 104407
97 euSerGluProLeuAsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAla 113
|||||
104408 TGTCTGAACCACTGGTGGGAACTTTTGGAGAAATACACCCAGCTTGC 104457
113 .....
104458 AGGTATCAGGAAATAGCGAGGAGAGTAGAATCTTTGTTGACAGTGT 104507
113 .....
104508 CCCTGGGTTCCAGATATGAGGGTAGAGCCTTGGAAGTCAGTGAAGAGT 104557
114 .....ArgIleuCysGlyIleThrPleuSerLeuGly 123
```

```
104558 TGTCAGTGTCCCTTCCCCCAGGGAATGTGACATCTGTGCTTCCCTGGGT 104607
124 GlyPheHisGluArgGlyGlnAspTrpGluIleAsnGlnLysIleTyrAs 140
|||||
104608 GGTTCATATGAGCGTGGCCAGAGCTGGGAGCAGACTCAGAAATCTACAA 104657
140 nCysHisValLeuLeuAsnSerLysGly..... 149
|||||
104658 TTGTACAGTGTGCTGCTGAACAGCAAGGTGAGACTTTAATACCTTTAGCC 104707
149 .....
104708 TGCCCTTCCCATGCTCTTCTACCTAGATTCTCCAGAAATGTTTCTCAAC 104757
149 .....
104758 TCTTATTCTTGTACCCAGAGATTAGGGGTGTCTACTTCACTTCCCTA 104807
149 .....
104808 GCCTATAAAGTATCTCCTCTGGGAGGAGTAACAGGCTTTAGAACAA 104857
150 .....Ser...ValValAlaSerT 155
|||||
104858 CCAGCACTGATATTCTTCTTCTTCTTACTGTAGGGGAGTAGTGCCCACTT 104907
155 YrArgLysThrHisLeuGlyAspValGluIleProGlyGlnIleProMet 171
|||||
104908 ACAGAAAGACATCTGTGTGACGTAGAGATCCAGGCGAGGGGCTTATG 104957
172 ArgGluSerAsnTyrThrLysProGlyGlyThrLeuGluProProVal 188
|||||
104958 TGTGAAGCAACCTCTACCATGCTGGGCCAGCTTTGATGTCACCTGTGAC 105007
188 sThrProAlaGlyLysValGlyLeu..... 196
|||||
105008 CACACACAGACAGGAGGTAGAGTGTGAAGAGATGAGGAGGGGAACAG 105057
196 .....
105058 GAATACTTTGAACAGCGCATAGAGATAGAAAGCCCTAAGAGAGGGGTA 105107
196 .....
105108 ATGGAATATGACTAGATGCTGTGACAACAGACAGAGAACTACTTAAG 105157
197 .....AlaIleCysTyrAspMetArg 203
|||||
105158 TAGGCTGTTTTCAATTCACAGATTGGTCTAGCTGTCTGTATGACATCGG 105207
204 PheProGluLeuSerLeuLysLeuAlaGlnAlaGluIleLeuThr 220
|||||
105208 TTCCCTTAACCTCTCTGTGGCATTTGGCTCAAGCTGGACAGAGATACTTAC 105257
220 rTyrProSerAlaPheGlySerValThrGlyProAlaHisTrpIleVal 236
|||||
105258 CTATCCCTTACGCTTTGTGATTCATTAACAGGCCAGCCACAGTGGGAGTAA 105307
237 Leu..... 237
105308 GATGATGCCCTTTTAAACATTAAGGGCTTTTCTTAACCTCACTTCCCC 105357
237 .....
105358 CTTTGGCCCTTACAGTTAAATTCCTTCCCTTTCACCTAATGGGAAAC 105407
237 .....
105408 TCATTTCCCAAGATATTCTCTCATGATAGTTAAATAGTAATCATTTCC 105457
237 ..... 237
```

```
105458 TAGCAATTACCAAAATAGTACAATGGGTAGATGCTGTAAATGCC 105507
237 ..... 237
105508 TCACCTGTACATCCCACTATTGCTACATGTAAGTAACACACA 105557
238 ..... LeuArgAlaArgAlaIleGluSerGlnCys 248
105558 TCTCATGCCAGGTGTGCTGCGGGCCCTGCTATGAAACCCAGTGCTA 105607
248 rValIleAlaAlaIleGlnCysGlyArgHisHisGluThrArgAlaSer 265
105608 TGTAGAGCAGCAGCAGCAGTGTGAGCCACCATGAGAGAGACAAGT 105657
265 yrcGlyHisSerMetValValAlaAspProTyrGlyThrValAlaAlaGly 281
105658 ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGTGGCCCTG 105707
282 SerGluGlyProGlyLeuGlyLeuAlaArgIleAspLeuHisPheLeuG 298
105708 TCTGAGGGGCGCAGGCTCTGCTTGCCTTGCCTGCAATAGACCTCACTATGCG 105757
298 nGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAspLeu 315
105758 ACAGTTGGCCGACACACCTGCTGTGTTCCAGCACGCGAGGCTGACCTG 105807
315 yrcGlySerLeuGlyHisProLeuSer 323
105808 ATGGCAATCTGGGTCAACCCACCTGCT 105833
seq_name: gb_hlg:AL354714_5
seq_documentation_block:
WPCOMMENT
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_0 1 110000
AL354714_1 100001 210000
AL354714_2 200001 310000
AL354714_3 300001 410000
AL354714_4 400001 510000
AL354714_5 500001 610000
AL354714_6 600001 699820
Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome 1)
Alignment_scores:
Quality: 1211.50 Length: 660
Ratio: 3.946 Gaps: 10
Percent Similarity: 46.515 Percent Identity: 41.667
Alignment_block:
US-09-357-675c-22 x AL354714_5 ..
Align seg 1/1 to: AL354714_5 from: 1 to: 110000
2 LeuGlyPheIleThrArgProHisGln.....LeuLeuGlySer 15
|||||..... 11
3858 CTGGGCTTCATCACCGAGGCTCTCACAGATTCCTGCTCTGTGTCTC 3907
15 rGlyTyrArgLeuLeuArgIleProValLeuGlySerHisGlnPro..... 29
|||||..... 11
3908 TGGACTCCGATACCTCAACTCACTACTTGTGTCTGCTCAGCCAGGTAA 3957
29 ..... 29
3958 ACCTTTTCTGTCTCTCAGTGCCTGCGCACTTATGATGCTCAGTTGTAA 4007
29 ..... 29
4008 ATGATAGTGGAGACACAGAGAGTGTCAATATCCACATTTGATTTGT 4057
29 ..... 29
4058 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGACATGATCAAA 4107
29 ..... 29
4108 GAATCCAGCTTTCCTGCTCTCCACTTGCACCCCTTAGATTAATTTG 4157
30 ..... ArgPr 31
4158 CTTCCTGTGCTATGAATCTGAGAAATCTGCTATGCTGTTCACAGCC 4207
31 oArgThrMetSer...SerSerThrSerTyrGluLeuProLeuValAla 47
|||||..... 11
4208 CAGAGCCATGCTATCTCTCTCTCTCTCTCTGGAACCTGCCCTGTGCTG 4257
47 aLcysGlnValThrSerThrProAsnLysGlnGluAsnPheLysThrCys 63
|||||..... 11
4258 TGTGCCAGTAACATCGACGACAGCAAGCAAGCAAGCAAGCAAGT 4307
64 AlAGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80
|||||..... 11
4308 GCTGAGCTGTGTGAGAGGCTGCCAGACTGGTCTGCTGCTGCTGCTTCT 4357
80 uProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 97
|||||..... 11
4358 GCCTGAGGCACTTGTACTTCACTGACGAGGACCTGCAAGAGCTTACACC 4407
97 eUeSerGluProLeuAsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAla 113
|||||..... 11
4408 TGTGTGAACACACTGGTGGGAACCTTTTGAAGAAATACACCACTTGCC 4457
113 ..... 113
4458 AGGTATCAGGAAATATACGAGAGAGTAGAATCTTGTGTGACAGTGT 4507
113 ..... 113
4508 CCTCGGCTTCCAGATATGAGAGGTAGAGCCCTTGGAAGTCACTAGAC 4557
114 ..... ArgGluCysGlyIleTyrPheSerLeuGly 123
4558 TGTACAGTGTCTTCCCTCCCGCAGGAATGTGACCTGCTGCTGCTGCT 4607
124 GlyPheHisGluArgGlyGlnAspTyrPheGluGlnAsnGlnLysIleTyr 140
|||||..... 11
4608 GATTTCATAGCGGTGGCCAAAGACTGGAGAGACTAGCAAAATCTACAA 4657
140 nCysHisValLeuLeuAsnSerLysGly..... 149
4658 TTGTACAGTGTCTGTGAAACGAAAGGTGAGACTTTTATACCTTTAGCC 4707
149 ..... 149
4708 TGCCTTCCCATGCTCTTCTACCTAGATTCTCCAGATTTCTCAAC 4757
149 ..... 149
4758 TCTTATTTCTTGACCCAGAGATTAGGGGTGTCTACTTCACTTCTCTA 4807
149 ..... 149
4808 GCCTATTAATATCTCTCTCTGAGAGAGAGTAGCAAGGCTTCTAGACA 4857
150 ..... Ser...ValValAlaSer 155
4858 CCAGCACTGATATTCTTCTTCTTACTGTAAGGGGAGTAGTGCCACT 4907
155 yArgGlyThrHisLeuGlyAspValAlaGluIleProGlyGlnGlyPromet 171
|||||..... 11
4908 ACAGGAAGACACATCTGTGTAGCTAGATTCACAGGCGAGGCGCTATG 4957
172 ArgGluSerAsnTyrThrLysProGlyGlyThrLeuGluProProValL 188
|||||..... 11
```

```
4958 TGTGAAGCAACTCTACCATGCTGGGCCAGTCTTGATGACCTGTGAG 5007
188 sThPrOlaIaGlyValGly.Leu..... 196
5008 CACACCCAGCAGGAGTAGAGTTGTGAAGGATGAGGAGGAGACAG 5057
196 ..... 196
5058 GAATACCTTGAAGTGGCAGTAGAGATGAAAGCCCTAAGAGAGGGGTA 5107
196 ..... 196
5108 ATGGAATATGACTAGATGCTGTGCAAAACAGAGCAGGAGACTACTAAG 5157
197 .....AlaIeCysTyrAspMetArg 203
5158 TAGCGTGTTCATTCAGATTGGTCTAGCTGTCTGCTATGACATGGGG 5207
204 PheProGluLeuSerLeuAlaGlnAlaGlyAlaGluLeuLeu 220
5208 TTCCTGAACTCTCTGTCAGATTGGCTCAAGCTGAGCAGAGATCTTAC 5257
220 rTyProSerAlaPheGlySerValThrGlyProAlaHisTrpGluVal. 236
5258 CTATCCTTTCAGCTTTTGGATCATTCATACAGCCACGCCACTGGGAGTAA 5307
237 .Leu..... 237
5308 GATGATGCTTTTAAACATAGAGGCTTTTCTTAACCTACTTCCCC 5357
237 ..... 237
5358 CCTGGCCCTACCAAGTTAAATTCCTCCCTTCCACCTAATGGGAAAC 5407
237 ..... 237
5408 TCATTCGCCAGATATTTCTCTCATGAAATAGTAAATAGTAATCATTC 5457
237 ..... 237
5458 TAGCAATTACCAAAATAGTCACAATGGGTAGATTGCTGTAATGTCCC 5507
237 ..... 237
5508 TCACCTGTCACTCCCACTATTGGCTACATGTACTTAAGTAAGACACA 5557
238 .....LeuArgAlaArgAlaIleGluSerGlnCysTy 248
5558 TCTCATGCCCGAGGTGTGCTGCGGGCCGCTCTATCGAAACCACTGCTA 5607
248 rValIleAlaIleAlaGlnCysGlyArgHisHisGluThrArgAlaSerT 265
5608 TGTAGTGGCAGCAGCAGTGTGACGCCCATGAGAAAGAGACAGCACTT 5657
265 rGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 281
5658 ATGCCACAGCATGTGTGTAGACCCCTGGGAAAGTGTGGCCCGCTGC 5707
282 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuG 298
5708 TCTTAGGGGGCAGGCTCTGCTTGCCTGCCGAATAGACCTCACTATCTG 5757
298 nGluMetArgIleHisLeuProValPheGlnHisArgArgProAspLeuT 315
5758 ACAATTGGCGCAGACACTGCTGTGTCTCAGACACCGCAGGCTGACCT 5807
315 rGlySerLeuGlyHisProLeuSer 323
5808 ATGGCAATCTGGTCACCCACTGTCT 5833
seq_name: gb_htg:AL590651
seq_documentation_block:
```

```
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
ACCESSION AL590651
VERSION AL590651.4 GI:13990236
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 167863)
REFERENCE
AUTHORS Harrison,E.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CH10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonequests@sanger.ac.uk
On May 7, 2001 this sequence version replaced gi:13751001.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA137A12
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 164729 bases at least Q40
Consensus quality: 165486 bases at least Q30
Consensus quality: 166169 bases at least Q20
Insert size: 167063; sum-of-contigs
Insert size: 171878; 3.4% error; agarose-fp
Quality coverage: 8.28x in Q20 bases; sum-of-contigs quality
coverage: 8.11x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 9373: contig of 9373 bp in length
* 9374 9473: gap of 100 bp
* 9474 29891: contig of 20418 bp in length
* 29892 29991: gap of 100 bp
* 29992 36239: contig of 6248 bp in length
* 36240 36339: gap of 100 bp
* 36340 103784: contig of 67445 bp in length
* 103785 103884: gap of 100 bp
* 103885 114878: contig of 10994 bp in length
* 114879 114978: gap of 100 bp
* 114979 126347: contig of 11369 bp in length
* 126348 126447: gap of 100 bp
* 126448 137764: contig of 11317 bp in length
* 137765 137864: gap of 100 bp
* 137865 163160: contig of 25296 bp in length
* 163161 163260: gap of 100 bp
* 163261 167863: contig of 4603 bp in length.
*
Location/Qualifiers
1. 167863
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-137A12"
/location="1p"
1. 9373
/note="assembly_fragment:02360
fragment_chain:1
clone_end:T7
vector_side:left"
```



```
misc_feature      9474..29891
/note="assembly_fragment:01854
fragment_chain:1"
misc_feature      29932..36239
/note="assembly_fragment:00805
fragment_chain:1"
misc_feature      36340..103784
/note="assembly_fragment:01467
fragment_chain:1"
misc_feature      103885..114878
/note="assembly_fragment:00991
fragment_chain:2"
misc_feature      114979..126347
/note="assembly_fragment:01503
fragment_chain:2"
misc_feature      126448..137764
/note="assembly_fragment:01297
fragment_chain:2"
misc_feature      137865..163160
/note="assembly_fragment:00117
fragment_chain:2"
misc_feature      163261..167863
/note="assembly_fragment:01267
fragment_chain:2
clone_end:SP6
vector_side:right"
BASE COUNT      43927 a 39935 c 40002 g 43197 t      802 others
ORIGIN

alignment_scores:
    quality: 1211.50      length: 660
    ratio: 3.946      gaps: 10
    percent similarity: 46.515      percent identity: 41.667

alignment_block:
US-09-357-675c-22 x AL590651/rev ..

Align seg 1/1 to reverse of: AL590651 from: 1 to: 167863

      2 LeuGlyPheIleThraGProPheIsgIn.....LeuLeuGlysrh 15
      |||||
151392 CTGGGCTTCATCAGCAGGCTCCTCTCAGATTCCTGCTTCCTGTGTC 151343
      15 rGlyTyraGleuLeuArgIleProValIleuGlysrhInPro..... 29
      |||
151342 TGGACTCCGGAATACCTCAACTCTCACTTGTGCTCAGCCAGGTAA 151293
      29 ..... 29
151292 ACCTTTTGTGTGCTCCTCAGTGCCTGGCACTTAGATGCTCAGTTTGTAA 151243
      29 ..... 29
151242 ATGGATAGTGGAGACACAGAGAGTGTCACTATCAACACATTGATTGCT 151193
      29 ..... 29
151192 GAGCCCTACTAGCCCTGGGTCAACGTGCTGTAGAGCATGTCAAAAG 151143
      29 ..... 29
151142 GAAGTCCAGCTTTCCTGCTCTCCACTTGACCCCTTAGCATTAATTG 151093
      30 .....ArgPr 31
151092 CTTCCTGTGCTATGAATCTGGAATTCCTGCTATGCTGTTCACAGGCC 151043
      31 oArgThrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 47
      |||||
151042 CAGAGCCATGAGCTATCTCTTCCTTCCTGCGAAGTCCCTGCTGTG 150993
      47 aLcysGlnValThrSerThrProAsnLysGlnLusnPhelyThrCys 63
```

```
|||||
150992 TGTGCCAGGTAACTGCAGCCAGACAAAGCAAGCACTTTAAAAACATG 150943
      64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80
      |||||
150942 GCTGAGCTGGTTCGAGAGGCTGCAGACTGGGTCTGCTCCCTGCTTCT 150893
      80 uProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 97
      |||||
150892 GCCTGAGGCAATTGACTTCATTGCAGCGGACCTGCAGAGAGCTTACACC 150843
      97 euSerGluProLeuAsnGlyAspLeuGlyGlnTrpSerGluLeuAla 113
      |||||
150842 TGTGTGAACCACTGGTGGGAACTTTTGGAGATACACCCACTTGCC 150793
      113 ..... 113
150792 AGGTATCAGGGAATATAGCAGAGAGGTAGAAATCTTTGTGACAGTGT 150743
      113 ..... 113
150742 CCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAGTCAAGAGT 150693
      114 .....ArgLcysGlyIleTrpLeuSerLeuGly 123
      |||||
150692 TGTCAAGTCCCTTCCCCCAGGGAATGTGACTCTGCTCTCTTGCGGT 150643
      124 GlyPheHisGluArgGlyGlnAspTrpGluGlnAsnGlnLysIleTyra 140
      |||||
150642 GGTTCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAAATCTACAA 150593
      140 nCysHisValLeuLeuAsnSerLysGly..... 149
      |||||
150592 TTTGCACGTGTGCTGTAACAGCAAGGTAGAGACTTATTAACCTTACCC 150543
      149 ..... 149
150542 TGCTCTTCCATCTCTTCTACTAGATTCTCCAGAAATGTTTCTCAAC 150493
      149 ..... 149
150492 TCTATTTCCTTGACCAAGGATTTAGGGGTGTCCTACTTCAGTTCTTA 150443
      149 ..... 149
150442 GCCTATTAACATATCTCTCTTGGAGAGAGTAAGCAAGCCTTCAAGACA 150393
      150 .....Ser...ValValAlaSerT 155
150392 CCAGACTGATATTCCTTCTTCTTACTGTAGGGGCACTAGTGGCCACTT 150343
      155 yTrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 171
      |||||
150342 ACAGAGAGACACATCTGTGTACGTAGAGATTCCAGGGCAGGGCCATATG 150293
      172 ArgGluSerAsnTrpThrLysProGlyGlyThrLeuGluProProVal 188
      |||||
150292 TGTGAAAGCAACTCATCCATCCCTGGGCCCGAGTCTGTAGTCACTGTGAG 150243
      188 sThrProAlaGlyLysValGlyLeu..... 196
      |||||
150242 CACACACAGCAGCAAGGTAGGATTTGTAAGAGATGAGGGAGGGAGCAAG 150193
      196 ..... 196
150192 GAATACTTTGAACAGTGCAGTAGAGATAGAAAGCCCTAAGAGAGGGGTA 150143
      196 ..... 196
150142 ATGGAATATCAGTAGATGCTGTGACAAACAGACAGACAGACTACTAAG 150093
      197 .....AlaIleCysTyraPmetArg 203
      |||||
```

150092 TAGGGCTTTTCATTCAGATTGGTCTAGCTGTCTGATGATGACG 150043  
 204 PheProGluLeuSerLeuLysLeuAlaGlnAlaGluLeuLeu 220  
 |||||  
 150042 TTCCCTGAACCTCTCTGTCATTTGGCTCAAGCTGGAGACAGATACTTAC 149993  
 220 rTyrProSerAlaPheGlySerValThrGlyProAlaHisTrpGluVal 236  
 |||||  
 149992 CTAATCCTTCAGCTTTTGATTCATTTACAGCCGACGCCACTGGAGGTAA 149943  
 237 .Leu..... 237  
 149942 GATGATGCTTTTAAACATAAGGCGCTTTCTTAACCTCATCTTCCC 149893  
 237 ..... 237  
 149892 CCTTGCCCTACCAGTTAAATTCCTTCCCTTCCACCTAAATGGGAAAC 149843  
 237 ..... 237  
 149842 TCATTCCCGAGATATTTCTCATGAAATAGTTAAATAGTAATCATTC 149793  
 237 ..... 237  
 149792 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGCC 149743  
 237 ..... 237  
 149742 TCACCTGTCACTTCCCACTATTTGCTACATGTACTTAAGTACACACA 149693  
 238 ..... 238  
 149692 TCTCATGCCAGAGTGTCTGCGGGCCGCTGCTATGAAACCAAGGCTA 149643  
 248 rValIleAlaAlaAlaGlnCysGlyArgHisHisGluThrAlaAlaSer 265  
 |||||  
 149642 TGTATGGGAGAGACACAGTGTGGAGCCACCATGGAAGAGCAAGTT 149593  
 265 YrGlyHisSerMetValValAspProTrpGlyThrValAlaAlaArgCys 281  
 |||||  
 149592 ATGGCCACAGCATGTGTGTAGACCTTGGGGAACAGTGTGGCCCTGTC 149543  
 282 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeu 298  
 |||||  
 149542 TCTGAGGGGCGAGGCTGTGCTGCGCCCAATAGACCTCAATATCTGG 149493  
 298 ngIlnMetArgGlnHisLeuProValPheGlnHisArgArgProAspLeu 315  
 :|||  
 149492 ACAGTGGCGCGACACCTGCTGTTCAGACACCGCAGGCGCTGACCTCT 149443  
 315 YrGlySerLeuGlyHisProLeuSer 323  
 |||||  
 149442 ATGGCAATCTGGTCAACCCACTGTCT 149417  
 seq\_name: gb-pr:AL591806  
 seq\_documentation\_block:  
 LOCUS AL591806 200822 bp DNA linear PRI 30-JAN-2002  
 DEFINITION Human DNA sequence from clone RP11-544M22 on chromosome 1, complete  
 sequence.  
 ACCESSION AL591806  
 VERSION AL591806.16 GI:18476709  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 AUTHORS Harrison, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT  
 On Feb 1, 2002 this sequence version replaced gi:17902927.  
 During sequence assembly data is compared from overlapping clones.  
 Where difference assemblies are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 chemistries were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em., EMBL; Sw.,  
 SwissProt; Tr., TrEMBL; Wp., WormPeP; information on the WormPeP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr1  
 RP11-544M22 is from the library RPCT-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6  
 This sequence is the entire insert of clone RP11-544M22 The true  
 left end of clone RP11-137A12 is at 156538 in this sequence. The  
 true right end of clone RP11-381D2 is at 145015 in this sequence.  
 Location/Qualifiers  
 1..200822  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-544M22"  
 /clone\_lib="RPCT-11.2"  
 10395..10397  
 /note="Sequence from overlapping clone RP11-381D2  
 (AL162592). Assembly confirmed by restriction digest."  
 44098..44107  
 /note="Sequence from overlapping clone RP11-381D2  
 (AL162592). Assembly confirmed by restriction digest."  
 59585..59604  
 /note="Sequence from overlapping clone RP11-381D2  
 (AL162592). Assembly confirmed by restriction digest."  
 90797..90883  
 /note="Single clone region. Reads generated from a  
 transposon library derived from a single pUC clone.  
 Restriction digest data confirm the assembly."  
 90861..90883  
 /note="Sequence from un1-directional dGTP big dye  
 terminator reads only."  
 BASE COUNT 55215 a 49052 c 46183 g 50372 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 1211.50 Length: 660  
 Ratio: 3.946 Gaps: 10  
 Percent Similarity: 46.515 Percent Identity: 41.667  
 alignment\_block:  
 US-09-357-675c-22 x AL591806 ..  
 Align seg 1/1 to: AL591806 from: 1 to: 200822  
 2 LeuGlyPheIleThrArgProHisGln.....LeuLeuGlyTrp 15  
 |||||  
 172789 CTGGGCTCAATCACAGGCTCTCTCACAGATTCCTCTGTCGTC 172838  
 15 rGlyTrArgLeuLeuArgIleProValLeuGlyTrpHisPro..... 29  
 ||| |||||





```
* 108600 111672: contig of 3073 bp in length
* 111773 111772: gap of unknown length
* 111773 114662: contig of 2890 bp in length
* 114663 114762: gap of unknown length
* 114763 116145: contig of 1383 bp in length
* 116146 116245: gap of unknown length
* 116246 118147: contig of 1902 bp in length
* 118148 118247: gap of unknown length
* 118248 120579: contig of 2332 bp in length
* 120580 120679: gap of unknown length
* 120680 122618: contig of 1939 bp in length
* 122619 122718: gap of unknown length
* 122719 124539: contig of 1821 bp in length
* 124540 124639: gap of unknown length
* 124640 126551: contig of 2012 bp in length
* 126552 126751: gap of unknown length
* 126752 128460: contig of 1709 bp in length
* 128461 128560: gap of unknown length
* 128561 130592: contig of 2032 bp in length
* 130593 130692: gap of unknown length
* 130693 132864: contig of 2172 bp in length
* 132865 132964: gap of unknown length
* 132965 134881: contig of 1917 bp in length
* 134882 136682: contig of 1701 bp in length
* 136683 136782: gap of unknown length
* 136783 139138: contig of 2356 bp in length
* 139139 139238: gap of unknown length
* 139239 140757: contig of 1519 bp in length
* 140758 140857: gap of unknown length
* 140858 142558: contig of 1701 bp in length
* 142559 142658: gap of unknown length
* 142659 144303: contig of 1645 bp in length
* 144304 144403: gap of unknown length
* 144404 145490: contig of 1087 bp in length
* 145491 145590: gap of unknown length
* 145591 147171: contig of 1581 bp in length
* 147172 147271: gap of unknown length
* 147272 148639: contig of 1368 bp in length
* 148640 148739: gap of unknown length
* 148740 150295: contig of 1556 bp in length
* 150296 150395: gap of unknown length
* 150396 151505: contig of 1110 bp in length
* 151506 151605: gap of unknown length
* 151606 152813: contig of 1208 bp in length
* 152814 152913: gap of unknown length
* 152914 155423: contig of 2510 bp in length
* 155424 155523: gap of unknown length
* 155524 156880: contig of 1357 bp in length
* 156881 156980: gap of unknown length
* 156981 158045: contig of 1065 bp in length
* 158046 158145: gap of unknown length
* 158146 159255: contig of 1110 bp in length
* 159256 159355: gap of unknown length
* 159356 160417: contig of 1062 bp in length
* 160418 160517: gap of unknown length
* 160518 161644: contig of 1127 bp in length
* 161645 161744: gap of unknown length
* 161745 163180: contig of 1436 bp in length
* 163181 163280: gap of unknown length
* 163281 164483: contig of 1203 bp in length
* 164484 164583: gap of unknown length
* 164584 166039: contig of 1456 bp in length
* 166040 167290: gap of unknown length
* 167290 167290: contig of 1151 bp in length
```

## alignment\_scores:

```
Quality: 1187.00      Length: 835
Ratio: 3.854          Gaps: 12
Percent Similarity: 36.886      Percent Identity: 35.569
```

## alignment\_block:

US-09-357-675C-22 x AC105589

```
Align seg 1/1 to: AC105589 from: 1 to: 178418
2 LeuGlyPheIIeThrArgProHISln.....LeuLeuCysTh 15
|||||
CTGGCTTCATCACCAGGCTTCATCATCCTGCTGCTGTGAC 39144
15 rGlyTyfArgLeuLeuArgIleProValLeuCysThrGlnPro..... 29
|||||
CGGATCCGATTACCTCAATCTCAGTACTTGTATCAGCCAGGTAAc 39194
29 ..... 29
39195 ATATTATTGTTCTTTGTCTAGCAGTAGTGTCTAATGCATGTTATTA 39244
29 ..... 29
39245 CCCGAATAGTGAGCTGCCATCCACCAAAATTAATGCTGTACTCCAG 39294
29 ..... 29
39295 TAGCTTAGGGTTATTAACTGCTGGAAGAAATGAACAGTCTAGGG 39344
29 ..... 29
39345 TTGGTGGACACACTCTAATCCACAGACTGGGAAACAGGAGAGAT 39394
29 ..... 29
39395 CAACTAGAGATTAAATCATATCCCGCTACATAGAAAGTCTGCTGGGC 39444
29 ..... 29
39445 TACAAAAGACCCGGCTCAAAAGGGAGCTAAATCAGAAAAAAG 39494
29 ..... 29
39495 TAACTTCGGCTCCACCCCACTGAGCTCCCAAGTTTCTCCGGGCA 39544
30 ..... ArgProArgTh 33
39545 CGCGGTAGAAATCTGAGAAATCTGCTGTATTCATTCACAGGCCAGAGC 39594
33 rMet.....SerSerSerThrSerTfPglLeuProLeuValAlaValc 48
|||||
CATGCGACAGTCTCATGACCTCTGAGAGCTGCCCTGGGCTGTGT 39644
48 ySGlnValThSerThrProAsnLysGlnGluAsnPheLysThrCysAla 64
|||||
GCCAGGTACATCAACACCAACAGCAGAGACATTAAACGTGTGCT 39694
65 GluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeu 81
|||||
GAGCTGTTCAAGAGCTACACGAGTACGAGTGGGGCTTCCCTGCTTGTGC 39744
39695 TGAGGCATTGCTTATTGACGAGAAATCTGCCAGACATATACCTGT 39794
81 oGluAlaPheAspPheIleAlaArgAsnProAlaLutThrLeuLeuLeu 98
|||||
erGluProLeuAsnGlyAspLeuLeuGlyIntYrSerGlnLeuAla... 113
|||||
CCGACCACTGGATGGGACCTTTTGGGCAATATATAGCCAGCTTCCAGG 39795
113 ..... 113
39845 TACAAGGTTAGGCTGGGGTGGGGGAGGAGGTANATCCCTGTGGGCG 39894
113 ..... 113
39895 TGTTCCTGATTCACCATAGAGGGGTAGAGTTGTCAGTGCTCCCTCC 39944
114 ..Arg.GluCysGlyIleTfPheSerLeuGlyGlyPheIleGluArgI 129
...:|||||
```

```

39945 CTCAGGGAAATGTGAATCTGGCTGCTCTGGGTGGTTCCACCACCAAGCTGG 39994
129 yglNapTPRgluInsnaglLysIleTyraSncYHisValLeuLeuA 146
| ::::::::::::::::::::|
39995 CCCAGATATGGGAACAGATCCGAAAAATCTACATTTGTCTGGGCTTCTGA 40044
146 sn.SerLys..... 148
| ::::::::::|
40045 ACAACCAGGGGTGAGACATTCATCTTCAACTTGGCTCTTCTTAAGGTC 40094
148 ..... 148
40095 ATCTAATCTCTCTCACTTAGAATTAACTTAATTTCTCACTGCTGGCT 40144
148 ..... 148
40145 TTAGGGGTTTTTTTGTGTTTGTGTAATGCTTCTTACGCTTAAAA 40194
148 ..... 148
40195 TTACTCTTTGGAGAGAGCTAGCTGACTTCAGATGCTGTACTGAGC 40244
149 ..... GlySerValValAlaSerTyraIglYstThiSL 160
40245 ATTTCTCTTACTTTTGGATCAGTACTGCGCAGTTACAGGAACACATC 40294
160 eucYasPvalIgluIleProglYngIyPrometaIrgIuseIasPntYr 176
| ::::::::::::::::::::|
40295 TATGTATGTAGATGATCCAGGTCAGGGGCTTATGAGAGAAACCACTAT 40344
177 ThLysProglYngIyThrLeuIgluProProValLysThrProAlaIglY 193
| ::::::::::::::::::::|
40345 ACCATCCTGAGATACGCTCTTGAAACACCTGTCAAGACACAGGACAA 40394
193 sValIglY..Leu..... 196
| ::::::::::|
40395 GGTAGGAGTTTGAAGAAATAAGATATGTTGAAAAAGTTGTAACATC 40444
196 ..... 196
40445 TCCCTTTGGTGTGATGATGAGTAGTATTCATGTTACTGTGACAG 40494
197 ..... AlaIle 198
40495 GACAGGAGATGATGAGTGGTGTCTCTCGTTTCAGGGTGGTCTAGCAATC 40544
199 CysTyraSPmeIarIghPheProglIuleuSerIleLysLeuAlaIgluAgl 215
| ::::::::::::::::::::|
40545 TGTATGACATCGGTCCCTGAACCTTCTCTGAAATGGGCTCAAGCTGG 40594
215 yAlaIgluIleuThrTyPProSerAlaIghIglYserValThrIglYProA 232
| ::::::::::::::::::::|
40595 GCCAGAAATACCTTACTTATCTTCACCTTCACCTTGTGATCTGTTACGGGCCAG 40644
232 IaHisTrpIgluVal..... 236
| ::::::::::|
40645 CCCACTGGAGGTAGAGAAATGCCCTCTTCATACATATAGGCTCCTTAA 40694
236 ..... 236
40695 ATCTATTTTGTAGTAGTGGGAGACAGATAGACAAACCTAAACTTAAGC 40744
236 ..... 236
40745 ACTTCAGGGGATAGCCACAGTTTCAAAGCCAGCATAGTATGCTCGAGT 40794
236 ..... 236
40795 AAAACAGAAACAAATFAGAGCTGGCAGTGGGGGTGCACACCTTTAAT 40844
236 ..... 236
40845 CCCAGCCTCAGAGAGCAGAAAGCAAGTGTGATGTGATCAAGACAGCCT 40894

```

```

236 ..... 236
40895 GGTCTACAGAGTAGTTCCAGAGACAGCGGCTACACAGAGAAACGGTC 40944
236 ..... 236
40945 TCAAAACAAACAAACAAACAAACAAAGTAACCCACACACCCAAA 40994
236 ..... 236
40995 ATAGTGTGATGATGGTGTGTGCTTTTATCTCCGACATGTAAGCAGA 41044
236 ..... 236
41045 GCGCGGACAGATCTCTGTGGGTCCAGGCCAACCAAGTTATACAAGAGAC 41094
236 ..... 236
41095 CTATATATAAAAAATACAAAACAAACACACCCCTCCCTCTACTTGA 41144
236 ..... 236
41145 AGATATTTCTTCCTCCGATCTGTGAGAGAAATTCGTTCTTAGATAGG 41194
236 ..... 236
41195 TGTGATGATGTTATAGTTAGTTACTTAATCTAGTTGATCATATAGTCAAT 41244
236 ..... 236
41245 GAGTAGATGGGCCACAGTGTCCGCTTCTGTCACTCGCCATTAATG 41294
237 ..... Leu..LeuArgAla 240
41295 CTGTATGTTCTTAGTAGACAGCATCTTCATCTCAGGTGCTGCGGGCC 41344
241 ArgAlaIleIgluSerIglYcysTyraValIleAlaIleAlaIgluIglY 257
| ::::::::::::::::::::|
41345 CGCGCATGTGATGCTCAGTCTGATGATGATGATGATGATGATGATGATG 41394
257 gHisHisIgluThrArgAlaSerTyrgIyHisSerMetValValAspProt 274
| ::::::::::::::::::::|
41395 CCACCATGAGACAAAGACAAAGTACCGCCACAGCATGTGTGTGACCCCT 41444
274 rPglYthrValValAlaIleArgCysSerIgluIleProglYleuCysLeuAla 290
| ::::::::::::::::::::|
41445 GGGGACAGTGTGTGGCCAGCTGTTCGAGGAGACAGGCTTTGCTTGTCT 41494
291 ArgIleAspLeuHisPheLeuIgluIleMetArgIleHisLeuProValph 307
| ::::::::::::::::::::|
41495 CGAATGATGATCCACATTCGCAACAGATGCGCAACACATTCGCTGTGT 41544
307 eGlnHisArgArgProAspLeuTyrgIySerLeuIgluHisProLeuSer 323
| ::::::::::::::::::::|
41545 CCAGCATCGCAGACCTGACCTTATAGCAGTCTC...CTGCCACTCTCC 41590
seq_name: gb_ov:AF284575
seq_documentation_block:
LOCUS AF284575 1214 bp mRNA linear VRT 22-JUL-2000
DEFINITION Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.
ACCESSION AF284575
VERSION AF284575.1 GI:9367117
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 1214)
Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
Bieganski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.

```

TITLE Crystal structure of the worm Nitfhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers

JOURNAL Curr. Biol. 10 (15), 907-917 (2000)

MEDLINE 20414396

REFERENCE 2 (bases 1 to 1214)

AUTHORS Pace, H.C., Hodawadekar, S.C., Draganescu, A., Huang, J., Bleganowski, P., Pekarsky, Y., Croce, C.M. and Brenner, C.

TITLE Direct Submission

JOURNAL Submitted (05-JUL-2000) Kimmel Cancer Center, Thomas Jefferson University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA

FEATURES

source Location/Qualifiers

1..1214

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

gene 1..1214

/gene="Nifl"

CDs 180..1046

/gene="Nifl"

/note="Nifl: similar to the Nit domains of Caenorhabditis elegans Nitfhit and Drosophila melanogaster Nitfhit"

/codon\_start=1

/product="Nit protein 1"

/protein\_id="AA87104.1"

/db\_xref="GI:9367118"

/translation="MAGAHKPLIAVCOMSTSDKKNFATCSRLIREAGRRACWFLPEAFDYIGSIEERLSIASLHPTIORYQLAREGIMLSLIGFHEKGPWPDORI SNGHYVDNTHGHIYSYRKALHPDVLQNGYSLESSSTLPGLIRITTPAKTIGLGVCTDLKPEPSIALAQOGALLTPPSAFITTLAHEVILRLRALETQCYVAAAO TDRHNERSTSGAMVYDPMGLVIGOCDEGTGYAEIDIPYMERVRDMVWRIRRT DLYGKISFNKPD"

BASE COUNT 281 a 305 c 345 g 282 t 1 others

ORIGIN

ALIGNMENT SCORES:

Quality: 933.50 Length: 311

Ratio: 3.734 Gaps: 2

Percent Similarity: 80.386 Percent Identity: 55.305

Alignment block:

US-09-357-675c-22 x AF284575 ..

Align seg 1/1 to: AF284575 from: 1 to: 1214

10 HisGlnLeuLeucThrglyTyraGleuLeu.....ArgIleProVa 24

11 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

102 CACAGCGCCCTTGTC.....AGACTGTATAGACTGTGACTTTAAC 142

24 LleucThrgInProArGProArGhrmetSerSerSerThrSerTrpG 41

13 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

143 CCTGTGTCGGATATAGGGGTTAGAGCCGTGAGTGCATGCGTGGTCCC 192

41 LuleuProLeuValAlaValAlaCysGlnValThrSerThrProAsnLysGln 57

193 ACAAGCCCTGATGTCGCGTGTGCACATGACTCAACCTCTGTATAGAG 242

58 GluAsnPhleLysThrcysAlaGluLeuValGlnGlnAlaAlaArgLeuG 74

243 AGAATTTCGCCACGCTGTCGGCGGTGATCCGGAGAGCTCGGGCGTCG 292

74 ValAcylsleuAlaPhleuProGlnAlaPhleuAspPheIleAlaArgAsn 91

293 CCTGTGATGAGGTGTTCTGCCGAGACCTTTGACTATATCGGGGGACGA 342

91 roAlaGlnThrLeuLeuLeuSerGlnProLeuAsnLysAspLeuGly 107

343 TTGAGAGAGACGCTGAGTCTGCTGACTCTCATATGGGACACACATTTCAG 392

108 GlnTySerGlnLeuAlaArgGlyCysGlyIleTrpLeuSerLeuGly 124

393 CGTTACACCACTCGCAGGGAGTGTGGCTGCTTCCCTCGGGGG 442

124 yPheHISGluArgGlyGlnAspTrpIleGlnAsnGlnLysIleTyraSnc 141

|||||..... ||| ::::: ::::: |||

443 ATTTCAATGAGAAAGAGACCAACTGGGACAGGACCAACGATTTCCAAAT 492

141 ySHISValLeuLeuAsnSerLysGlySerValValAlaSerTyraLys 157

493 CTCACCTGTTGGTGGACAAACAGGACATAGATGGTGTACCCGAG 542

158 ThrHisLeuCysAspValGluIleProGlyGlnGlyProMetArgGly 174

543 GCTCACCTGTTGACGTAGACTGCAGAAATGAGTGTCACTCAGAGAG 592

174 rAsnTyTrpLysProGlyGlyThrLeuGluProProValLysTrpPro 191

593 CAGTTCCACCTCCCGGACAGACGTTATTCGCCCATCACTCTCCAG 642

191 laGlyLysValGlyLeuAlaIleCysTyraSphetaArgPheProGluLeu 207

643 CAGAAAGATTGGCTGGGGGTGTGTACGACCTCCCTCCAGAAATTC 692

208 SerLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrosAl 224

693 TCTTGCTGTGGCCCAACAGAGGAGAACTTCTCACTTACCCTTCTGC 742

224 aPheGlySerValThrglyProAlaHisTrpGluValLeuLeuArgAla 241

743 CTTCACCCCTCACTACTGTCTGCGACATTTGGAGGTCTTGTGAGACC 792

241 rGlaIleGluSerGlnCysTyraValIleAlaAlaGlnCysGlyArg 257

793 GTCCCATAGAAACCACTGCTAGCTTACGTCAGCGCACAGACAGACA 842

258 HisHisGluThrArgAlaSerTyrglyHisSerMetValValaProTr 274

843 CACATGAGAGAGAGAGCTCTATGTCACGCTATGTGTGGTAGACCCGTG 892

274 rGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAla 291

893 GGGGCTGTGATTTGGCCATCCAGAGAACAGACAGATATGTATGCTG 942

291 rGleAspLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe 307

943 AGATTGACATTCCTCATATGAGAGGTGTGAGGGGAGCATGCCGTGTGG 992

308 GlnHisArgArgProAspLeuTyrglySerLeu 318

993 AGGCACCGCAGAGACTGTATGTGGAATAATC 1025

seq\_name: gb\_in:AF069986

seq\_documentation\_block: 1385 bp mRNA linear INV 23-JUL-1998

LOCUS AF069986

DEFINITION Caenorhabditis elegans nitrlase and fragile histidine triad fusion protein Nitfhit (Nifhit) mRNA, complete cds.

ACCESSION AF069986

VERSION AF069986.1 GI:3228663

KEYWORDS

SOURCE

ORGANISM

Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE

1 (bases 1 to 1385)

Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tiliib, S., Draganescu, A., Wernuth, P., Rothman, J.H., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

Nitrlase and fhit homologs are encoded as fusion proteins in drosophila melanogaster and caenorhabditis elegans

Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

JOURNAL 9837986

MEDLINE

REFERENCE

2 (bases 1 to 1385)

Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tiliib, S., Draganescu, A., Wernuth, P., Rothman, J., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

Direct Submission

TITLE

JOURNAL Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

FEATURES  
 source 1..1385  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 1..1385  
 /gene="Nlfhlt"  
 3..1325  
 /gene="Nlfhlt"  
 /codon\_start=1  
 /product="nitrilase and fragile histidine triad fusion protein Nlfhlt"  
 /protein\_id="AAC39136.1"  
 /db\_xref="GI:3228664"  
 /translation="MLSTVPRRTMATGRHFIVACOMTSNDLEKNPOAKNMTIERAGE KKEVMFLPECEPFELINKNEQIDLAMTDCDEMEKRELARKHNITLSISGLHNKP SDANHPWNTHLIIIDSGVTRAEYTKLHLEFLETPGKRLMESEFSKGTMTIPVDTP IGRGLSTICDVPRPELISLWNRKGAQLSPSPALFNLGALMETLLARALENDCY VVAAOCTGAHNPKEQSGHSMVDPNGAVVACSEKRVDMCFADLISYVDLEMOYV FSHRSDLYTLHNEKSESGELKFAFENIPADHIFSTPSFVFLKRVTDGHVLY SPKRVVPLDLDAEDLFIYAKKVOAMLEKHNHTSTTICVODGKDGQVPHNH IHLPRRADDFGNETYOKLASDKPERKPRSNEDMAEAVVYRNLM"

BASE COUNT 399 a 314 c 326 g 346 t

ORIGIN

alignment\_scores:  
 Quality: 706.00 Length: 284  
 Ratio: 3.284 Gaps: 2  
 Percent Similarity: 75.704 Percent Identity: 48.592

alignment\_block:  
 US-09-357-675C-22 x AF069986 ..

Align seg 1/1 to: AF069986 from: 1 to: 1385

```

32 ArgThmEtsrSerThsrTrpGlulLeuProLeuAlaAlaValCy 48
|||||
24 CGAACAAATGGGAGCTGACGCGCATTT.....ATCGCAGTATG 61
|||||
46 SGLValThsrThrProAsnLysGlnGluAsnPhelYsthrCysAlaG 65
|||||
62 TCAATATGACGACGCAACAGATCTGGAAGAAAGAACTTCAAGCCGCAAA 111
|||||
65 LuleValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhelPro 81
|||||
112 ATATGATCGAAGACCGGAGAAAGAAAGTGGAGATGTTCTTACCA 161
|||||
82 GluAlaPhasPheIleAlaArgAsnProAlaGluThrLeuLeuSe 98
|||||
162 GAATGCTTCATTCATGGGCTCAACAAATGAGCAATGATGTTGGC 211
|||||
98 rGlulProLeuAsnGlyAspLeuLeuGlyGlnTySerGlnLeuAlaArg 115
|||||
212 GATGCTACGACGCTGTAATATGGAATAATCGGAGATTGGCGCAAA 261
|||||
115 LucysGlyIleTrpLeuSerLeuGlyGlnPheHisGluArgGlyLinsp 131
|||||
262 AACACAAATATTTGGCTTTCACCTAGAGGCGCTTCATCACAG....GAT 305
|||||
132 TrpGluGlnAsnGlnLysIleTyAsnCysHisValLeuLeuAsnSerly 148
|||||
306 CCATCAGACGCTGCTCATCCATCGAACACCCATCGATCATAGATTCCGA 355
|||||
143 sGlySerValAlaAserTyArgLysThrHisLeuCysAspValGluI 165
|||||
356 TGGAGTACCTGCTGCTGAGTACAAACAGCTTCACCTTCATTTGGAGA 405
|||||
165 lPepGlyGlnGlyPrometArgGluSerAsnTyThrLysProGlyGly 181
|||||
406 TTCTGGAAAAGTACGATTGATGGAGCGAGTTCAGCAAGCTGGCACT 455

```

```

182 ThrLeuGluProProValLysThrProAlaGlyLysValGlyLeuAlaI 198
|||
456 GAGATGATTCCTCCAGATTGACACCCCTATCGGACGCTCGGACTGCAT 505
|||
198 eCysTyArgMetAlaArgPheProGluLeuSerLeuLeuAlaGlnAlaG 215
|||
506 TTGCTACGACGCTGCTGCTCCAGAGCTGCTTGGGAACCGCTAAACGG 555
|||
215 LValGluIleuThrTyProSerAlaPheGlySerValThrGlyPro 231
|||
556 GCGCCCACTGCTCATTTCCAAAGTCATTCACACTGATACAGCTT 605
|||
232 AlaHisTrpGluValLeuLeuAlaArgAlaAlaGluSerGlnCysTy 248
|||
606 GCTCATGGGAACCTGCTGCTGCGCCGCGCAATGAGATCAGAGTCA 655
|||
248 rAlaIleAlaAlaGlnCysGlyArgHisHisGluThrAlaGlyAsn 265
|||
656 TTGTGTGGCGGCTGCTCAACCTGGGCGCCCAATCCGAACAGACTCT 705
|||
265 YrGlyHisSerMetValAlaAspProTrpGlyThrValAlaAlaArgCys 281
|||
706 ATGGGCATTCGATGCTGTGATCCGTGGGCGGCTGTTCGACAGTGT 755
|||
282 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuG 298
|||
756 TCCGAGAGAGTTCACATGCTGTGCTGTAATGAGACTGAGCTAGCTGA 805
|||
298 nGlnMetArgGlnHisLeuProValPheGlnHisArgAlaGlyProAsp 315
|||
806 CACTCTCTGTGAATGCAACACAGTATTCGCCACAGACGCTCGATCTCT 855
|||
315 Yr 315
|||
856 AC 857

```

seq\_name: gb\_in:AF069989

seq\_documentation\_block:

LOCUS AF069989

DEFINITION Drosophila melanogaster nitrilase and fragile histidine triad

fusion Protein Nlfhlt (Nlfhlt) mRNA, complete cds.

ACCESSION AF069989

VERSION AF069989.1 GI:3228669

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

Location/Qualifiers  
 1..1521  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /map="61A"  
 1..1521  
 /gene="Nlfhlt"







OM of: US-09-357-675C-22 to: N\_Geneseq\_032802:\* out\_format : pfs  
Date: Apr 29, 2002 10:15 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Comugen Ltd.

Command line parameters:

```
-MODEL=frame+P2n.model -DEV=x1h  
-O/cgn2.1/USP10.spool/US09357675/runat_29042002.091757_3254/app-query.fasta_1.1208  
-DB=N_Geneseq_032802 -Qfmt=fastap -SUFFIX=P2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsnum62  
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=ptc  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09357675 @CGN1_1_412 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

Query: US-09-357-675C-22  
Query length: 323  
Database: N\_Geneseq\_032802:\*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 338.150000

score list:

Sequence	Strd Orig	ZScore	EScore	len	! Documentation
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +		1482.00	2562.06	2.3e-13	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ6257 +		1457.00	2518.70	6.0e-13	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ30398 +		965.00	1661.85	3.2e-84	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAZ54548 +		933.50	1606.80	3.8e-81	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI12225 +		699.50	1196.34	2.7e-5	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI12224 +		699.50	1187.08	9.0e-5	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI20222 +		699.50	1187.01	1.7e-5	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ84214 +		483.00	822.93	1.7e-3	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAZ54547 +		483.00	819.66	2.6e-37	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ81479 -		474.50	805.37	1.6e-36	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ15102 +		410.00	650.08	7.3e-28	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ21607 +		410.00	632.83	6.7e-27	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAZ1430 +		389.50	660.97	1.8e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI1378 +		389.00	647.35	1.0e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI1376 +		389.00	645.31	1.0e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ58825 +		369.00	524.98	6.8e-21	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ72312 +		298.00	501.57	1.4e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ71819 +		298.00	501.57	1.4e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ34737 +		292.50	488.98	6.9e-19	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ28550 -		289.50	436.07	6.1e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ2205 +		275.00	431.15	1.1e-15	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:AAZ30052 -		261.00	417.23	6.8e-15	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAZ74332 +		259.50	396.58	9.6e-14	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAZ82006 -		238.00	403.09	4.2e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAZ91910 +		238.00	403.09	4.2e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ52691 +		237.00	396.23	1.0e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ54332 +		237.00	386.02	8.1e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ41226 -		234.50	376.66	7.6e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ41341 +		225.00	376.03	1.3e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ54306 +		219.00	350.19	3.7e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ2154 +		210.00	350.74	1.2e-11	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ6975 +		210.00	340.46	1.3e-10	

/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ6965 + 208.00 336.97 2.0e  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ6973 + 206.00 333.48 3.2e  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ6966 + 206.00 333.48 3.2e  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ6968 + 206.00 333.48 3.2e  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ6967 + 206.00 333.48 3.2e

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101

seq\_documentation\_block:

ID AAZ46101 standard; DNA: 1416 BP.

AAZ46101;

05-MAY-2000 (first entry)

Nucleotide sequence of the coding region of NIT1 gene.

NIT1 gene; nitrlase; tumour suppressor gene; FHIT; chromosome 3p14.2;

FRA3B; cancer; genome allele inactivation; ss.

Homo sapiens.

Mus sp.

Drosophila melanogaster.

Caenorhabditis elegans.

Location/Qualifiers

3..1415

/\*tag= a

/product= (pos: 18..20, aa: Xaa)

/product= (pos: 1179..1181, aa: Gln)

/product= (pos: 1182..1184, aa: Ala)

/note= "contains 7 internal stop codons; Xaa is an

unspecified amino acid"

WC200003685-A2.

27-JAN-2000.

20-JUL-1999; 99MO-US16366.

20-JUL-1998; 98OS-0093350.

(UYJE-) UNIV JEFFERSON THOMAS.

Croce CM;

WPI; 2000-171195/15.

P-PsDB; AAZ68739.

Novel nitrlase homologs used as diagnostic and therapeutic reagents

for the detection and treatment of cancer -

Claim 6; Fig 6; 25pp; English.

The present sequence represents the coding region of human, murine,

Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The

mammalian gene family with homology to bacterial and plant nitrlases.

The tumour suppressor gene FHIT in D. melanogaster and C. elegans code

for fusion proteins in which the Fhit domain is fused with a Nit domain.

In mouse and humans, FHIT and NIT are encoded by two different genes,

localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The

human FHIT gene at chromosome 3p14.2, spanning the constitutive

chromosomal fragile site FRA3B, is often altered in most common forms

of human cancer. The Nit1 protein overcomes the mutated inactivation

of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives

and analogues of them, and antibodies are used as diagnostic and

therapeutic reagents for the detection and treatment of cancers.

Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;

alignment\_scores:

Quality: 1482.00 Length: 326  
Ratio: 4.827 Gaps: 2  
Percent Similarity: 94.172 Percent Identity: 84.356

alignment\_block:

US-09-357-675c-22 x AA246101 ..

Align seg 1/1 to: AA246101 from: 1 to: 1416

```

2 LeuGlyPheIleThrArgProProHisGln.....LeuLeucysm 15
|||||
111 CTGGGCTTCATCACAGGCGCTCTCCACAGATCCCTGCTCTGTGTC 160
15 rGlyTyrArgLeuLeuArgIleProValLeuCysThnGlnProArgPro 32
|||
161 TGGACATCCGATACCTCACTCACTGACTTTGTGCTCAGCCAGGCCCA 210
32 rGlyMetSer...SerSerThrSerTrpGluLeuProLeuValAlaVal 47
|||
211 GAGCCATGGCTATCTCTCTCTCTCCGACACTGCCCTGGGGCTGTG 260
48 CysGlnValThrSerThrProAsnLysGlnLysAsnPhelThrcysAl 64
|||
261 TGGCAGGTACATCGAGCCAGACAGAACAGAACTTTAAACATGTGC 310
64 agLLeuValGlnGlnAlaAlaArgLeuGlnYAlaCysLeuAlaPheLeu 81
|||||
311 TGGAGCTGTTCGAGAGGCTGCCAGACTGGGCTGCTGCTTCTCTGC 360
81 rGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 97
|||
361 CTGAGCATTTGACTTCATTCAGCAGGAGCCCTGCAGAGAGCTCAACCTG 410
98 SerGluProLeuAsnGlnAspLeuLeuGlyGlnTyrSerIleuAlaAr 114
|||||
411 TCTGAACCATGTGGTGGGAAACTTTGGAAGAAATACACCAGCTTGGCAG 460
114 gGluCysGlyIleTrpLeuSerLeuGlyGlyPheHisGlnArgGlyGln 131
|||||
461 GGAATGTGAGCTGTGGCTGTCTTGGTGGTTTCCATGAGCGTGGCCAG 510
511 ACTGGAGGAGCACTCAGAAATCTACAAATGTGCAGCTGTGGAACAGC 560
131 sPTTrpGluGlnAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 147
|||||
511 ACTGGAGGAGCACTCAGAAATCTACAAATGTGCAGCTGTGGAACAGC 560
148 LysGlySerIleValAlaSerTyrArgLysThrHisLeucysAspVal 164
|||||
561 AAAGGGGACATAGTGGCCACTTACAGAGAACACATCTGTGTACGCTAGA 610
164 uileProGlyGlnGlyProMetArgGlnSerAsnTyrThrLysProGly 181
|||||
612 GATTCACAGGCGAGGGGCTATGTGTGAAGCAACTCTACCATGCTTGGC 660
181 LyrThrLeuGluProProValLysThrProAlaGlyLysValGlyLeuAla 197
|||
661 CCAGTCTTGTAGTCACTGTGCAGACACACAGCAGGCAAGATTGGTCTAG 710
198 IleCysTyrAspMetArgPheProGluLeuSerLeuLysLeuAlaGln 214
|||
711 GTCTGTATGACATGCGGTTCCCTGAACTCTCTGCAATGTGCTCAAGC 760
214 agLValGlnIleLeuThrTyrProSerAlaPheGlySerValThrGly 231
|||
761 TGGAGCAGAGATACTTACCTATCTTCAGCTTTGGATCCATTACAGGCC 810
231 roAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluSerGlnCys 247
|||
811 CAGCCCACTGGAGGTGTGCTCGGGGCCGCTGATCCGAACCCAGTGC 860
248 TyrValIleLeuAlaAlaGlnCysGlyArgHisHisGluThrArgAla 264
|||
861 TATGTAGTGGCAGCAGCAGTGTGAGCCGCCACCATGAGAAAGAGAGCAG 910
XX

```

```

264 rTyrGlnHisSerMetValValAspProTrpGlyThrValValAlaArgC 281
|||||
911 TATGTGCGCAGCATGTGTAGACCCCTGGGAGACAGTGTGGCCGCT 960
281 ysserGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeu 297
|||||
961 GCYCTAGGGGGCCAGGCTCTGCTGGCCGATTAACCTCAACTATCTTG 1010
298 GlnGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAsp 314
|||
1011 CGACACTTGGCGCGACACCTGCTGTGTTCACAGCCAGGCGCTTACCT 1060
314 yTyrGlySerLeuGlyHisProLeuSer 323
|||||
1061 CTATGGCAATCTGGTACCCACTGTCT 1088
seq_name: /SIDS1/gcdata/hoid-geneseq/geneseq-emb1/NA2000.DAT:AAF16257
seq_documentation_block:
ID AAF16257 standard; cDNA; 1382 BP.
XX
AC AAF16257;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytoskeletal; cardioactive; immunomodulatory; muscular;
KW vulnerrary; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
XX Homo sapiens.
OS
XX WO20005174-A1.
PN
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05988.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587513/55.
XX P-PSDB; AAB57054.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
XX cancer antigens, useful for treatment, prevention, and diagnosis of
XX disorders such as prostate cancer -
XX
XX Claim 1; Page 1124; 2338pp; English.
XX
XX AAF15566 to AAF15505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytoskeletal,
XX cardioactive, immunomodulatory, muscular, vulnerrary, gastrointestinal,
XX nephrotoxic, antiinfective, gynaecological and antibacterial activities,
XX and can be used in gene therapy. The prostate cancer antigen
XX polynucleotides may be used for detection of prostate cancer, chromosome
XX identification, as chromosome markers, and for numerous other diagnostic
XX or research purposes. The prostate cancer antigens may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wounds, and infectious diseases. AAF15506 to AAF15514 to
XX AAB57303 represent sequences used in the exemplification of the present
XX invention.
XX

```

50 Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

# alignment\_scores:

Quality: 1457.00 Length: 327  
Ratio: 4.746 Gaps: 2  
Percent Similarity: 93.884 Percent Identity: 84.098

# alignment\_block:

US-09-357-675c-22 x AAF16257 ..

Align seg 1/1 to: AAF16257 from: 1 to: 1382

```

1 MetLeuGlyPheIleThrArgProProHisGln.....LeuLeuCy 14
|||||
77 ATGCTGGGCTTCATCCACGAGCCCTCCACAGATTCCTGTCCTCTGCTG 126
14 SThrGlyTyrArgLeuLeuArgIleProValIleuCySThrGlnProArg 31
|||||
127 TCCGAGACTCCGGATACCTCACTCACTCACTTGTGCTCAGGCCAGGC 176
31 roArgThrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 46
|||||
177 CCAAGAGCATGGCTATCTCTCTCTCTCTGCAACTGCCCTGGTGACT 226
47 ValCysGlnValIleThrSerThrProAsnLysGlnLysAsnPhelySthrcy 63
|||||
227 GTGGCCAGGTAAACATCAGCCGCCAGACMACACAGAACTTAAACATG 276
63 sAlaGluLeuValGlnGlnAlaAlaArgLeuGlyAlaCysLeuAlaPheL 80
|||||
277 TGCCTGACCTGGTTGAGAGAGCTGCCAGACTGGGCTGCTGCTGCTTCC 326
80 eupProGlnAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 96
|||||
327 TGCCTGAGCATTTGACTTCATTCACAGGAGCCTGCAGAGACGCTACAC 376
97 LeuSerLupProLeuAsnGlyAspLeuLeuGlyLntIlySerGlnLeuAl 113
|||||
377 CTGTGAACCCATCGGTGGTGGAACTTTGGAAATACACCCAGCTTGC 426
113 aArgGlyCysGlyIleThrLeuSerLeuGlyGlyPheHisGlnArgIlyG 130
|||||
427 CAGGGATGTGGACTCGCTGCTCTGCTGCTGCTTCCATAGCGTGGCC 476
130 lAspTrpGluGlnAsnGlnLysIleTyrAsnCySHisValLeuLeuAsn 146
|||||
477 AAGACTGGAGACAGACTCAGAAATCTACAATGTCTCAGCTGCTGAC 526
147 SerLysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspVa 163
|||||
527 AGCAAAAGGGCAGTAGTGGCCACTTACAGGAACACATCTGTGTGAGCT 576
163 lGluIleProGlyLysGlnLysProMetArgGluSerAsnTyrThrLysPro 180
|||||
577 AGAGATTCCAGGCGAGGGGCT.ATGTGGAAGACAACTCTACCATGCTG 625
180 llyGlyThrLeuGluProProValIlyThrProAlaGlyLysValGlyLeu 196
|||||
626 GGCCAGACTCTTGAAGTCACTGTACACACAGCAGCAGGCAAGATTGCTCA 675
197 AlAlaIleGlyTyrAspMetArgPheProGluLeuSerLeuLysLeuAla 213
|||||
676 GCTGTCTGTATGACATGCGGCTCCCTGAACCTCTCTGCGGATGGGCA 725
213 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerValIlyThr 230
|||||
726 AGCTGGAGACAGATACCTTACCTTCACTTTCGATTTCATTCATACAG 775
230 lProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluSerGln 246
|||||
776 GCCCAGGCCACTGGAGGAGGTGTTGCTGCGGGCCCGGTATGAAACCCAG 825

```

```

247 CysTyrValIleAlaAlaAlaGlnCysGlyYArgHisIleGluThrArgAl 263
|||||
826 TGTATGTAGTGGCAGCAGCAGCAGCTGTGAGACGCCACCATGAGAGAGAGC 875
263 aSerTyrGlyHisSerMetValValaAspProTrpGlyThrValValAla 280
|||||
876 AAGTTATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGATGTGGCCCC 925
280 rGlySerGlyGlyProGlyLeuCySLeuAlaArgIleAspLeuHisPhe 296
|||||
926 GCTGCTCTGAGGGGCGCAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 975
297 LeuGlnIleMetArgGlnHisLeuProValPheGlnHisArgArgProAs 313
|||||
976 CTGGCAGAGTGGCGCGCAGACACTGCTGTGTTCAGCAGCAGCAGGCTCA 1025
313 pLeuTyrGlySerLeuGlnHisProLeuSer 323
|||||
1026 CCTCTATGGCAATCTGGGTGACCCACTGTCT 1056
seq_name: /STDS1/gcgdata/hold-geneseg/genesegn-emb1/NA1999.DAT.AAX30398
seq_documentation_block:
ID AAX30398 standard; DNA; 1203 BP.
XX
AC AAX30398;
XX
DT 14-MAY-1999 (first entry)
XX
DE DNA encoding a human secreted protein.
XX
KW Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS Homo sapiens.
XX
PN W0907891-A1.
XX
PD 18-FEB-1999.
XX
PE 04-AUG-1998; 98MO-US16235.
XX
PR 19-AUG-1997; 97US-0056732.
XX
PR 05-AUG-1997; 97US-0054798.
XX
PR 05-AUG-1997; 97US-0054803.
XX
PR 05-AUG-1997; 97US-0054804.
XX
PR 05-AUG-1997; 97US-0054806.
XX
PR 05-AUG-1997; 97US-0054807.
XX
PR 05-AUG-1997; 97US-0054808.
XX
PR 05-AUG-1997; 97US-0054809.
XX
PR 05-AUG-1997; 97US-0054809.
XX
PR 05-AUG-1997; 97US-0055309.
XX
PR 05-AUG-1997; 97US-0055310.
XX
PR 05-AUG-1997; 97US-0055312.
XX
PR 05-AUG-1997; 97US-0055386.
XX
PR 05-AUG-1997; 97US-0055311.
XX
PR 18-AUG-1997; 97US-0055970.
XX
PR 18-AUG-1997; 97US-0055986.
XX
PR 19-AUG-1997; 97US-0056365.
XX
PR 19-AUG-1997; 97US-0056366.
XX
PR 19-AUG-1997; 97US-0056557.
XX
PR 19-AUG-1997; 97US-0056370.
XX
PR 19-AUG-1997; 97US-0056371.
XX
PR 19-AUG-1997; 97US-0056563.
XX
PR 19-AUG-1997; 97US-0056731.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;

```



50 Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;

alignment\_scores:

Quality: 933.50 Length: 311  
Ratio: 3.734 Gaps: 2  
Percent Similarity: 80.386 Percent Identity: 55.305

alignment\_block:

US-09-357-675c-22 x AAD25458 ..

Align seg 1/1 to: AAD25458 from: 1 to: 1214

```

10 HisGlnleucysthrGlyTyrArgLeu.....ArgIleProva 24
102 CACACGCCCTGTGTGC.....AGACTGCTATAGACTGACTTAAAC 142
24 lleuCythrGlnProArgProArgThrMetSerSerSerThrSertrp 41
143 CCGTGTCCCGATATAGGGGTAGAACCTGAGTGCATATGGCTGTGCC 192
41 lueProleuValAlaValCysGlnValThrSerThrProAsnlyGln 57
193 ACAAGCCCTGATGCGCGTGCAGATGACTTCAACCTGTGATAGAGAG 242
58 GluAsnPhelyThrCysAlaGluLeuValGlnGluAlaAlaArgIle 74
243 AAGAAATTCGCCACGTTGCGCGCTGATCCGGAGGCTCGCGGCGTCG 292
74 yAlaCysleuAlaPheleuProGluAlaPheAspPheIleAlaArgAsp 91
293 CGCTTGATGTGTTCTGCGCGAGACCTTTGACTATATGGGGGACGA 342
91 roAlaGlnThrLeuLeuSerGluProleuAsnGlyAspLeuGly 107
343 TTGAGAGACGCTGAGTGTGCTGCTACATGGGAGACACATTACAG 392
108 GlnTyrSerIleuAlaArgGluCysGlyTyrIlePleuSerLeuGly 124
393 CGTTACACCCCACTCGCCAGAGTGGGCTCGGCTTCCCGGGGGG 442
124 yPheHisGluArgGlyGlnAspTrpGluGlnAsnGlnlyIleTyrAsn 141
443 ATTTCTATGAAAGAGACCACTGGACAGGACCAAGCATTTCCAAAT 492
141 yHisValleuLeuAsnSerlySeryValAlaIleSerTyrArglys 157
493 CTCACGGTGTGGACACACAGGCGACATATGATGGGTACGCGACAG 542
158 ThrHisLeuCysAspValGluIleProGlyGlnIleProMetArgIle 174
543 GCTCACCTGTTGACGTAGACTTCAGAAATGAGTGCATCAGACAGAG 592
174 rAsnTyrThrlyProGlyThrleuGluProProValIlyThrProa 191
593 CAGTTCACCCCTCCCGGACAGAGCTTATTCGCCCATCTCTCCAG 642
191 laGlyValAlaGlyLeuAlaIleCysTyrAspMetArgPheProGlu 207
643 CAGAAAGATGGGCTGGGGGTGTGTACGACCTCCGCTTCCGAAATTC 692
208 SerleuLysleuAlaGlnAlaGluIleLeuThrTyrProSerAl 224
693 TCCGTGCTGTGGCCCAAGAGAGCAACTTCTCACTTACCTCTGTGC 742
224 apheGlySerValThrGlyProAlaHisTrpGluValleuLeuArgAla 241
743 CTTCACCTCTACTGTGTCTGGACATGTGGAGGTGTGTGTGAGAGCC 792
241 rGaIaIleGlnSerGlnCysTyrValIleAlaAlaIleCysGlyArg 257
793 GTGCGATAGAAACCACTGCTAGCTAGTGTGAGGCGACAGACAGACA 842

```

```

258 HisGlnThrArgAlaSerTyrGlnHisSerMetValAlaAspProtr 274
843 CACATGAGAAAGAGACGCTCTATGCTACGCTATGGGTAGACCCGTG 892
274 rGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysleuAla 291
893 GGGGCTGTCTATGGCCATGCGCAGAGACAGACGAAATATGTATGCTG 942
291 rGlyleAspLeuHisPheLeuGlnGlnMetArgGlnHisleuProvalPhe 307
943 AGATTGACATTCCTCATGATGAGCGTGTGAGCGGACATGCGGTGTGC 992
308 GlnHisArgArgProAspLeuTyrGlySerleu 318
993 AGCACCGCAGAGACTGATGTATGGAAATATC 1025

```

seq\_name: /sids1/gcgsdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL12225

seq\_documentation\_block:

ID ABL12225 standard; cDNA; 1495 BP.

ABL12225;

26-MAR-2002 (first entry)

Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ss.

Drosophila melanogaster.

WO2001/1042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

P-PSDB; ABB68122.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Claim 1; SEQ ID NO 31157; 21np + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;

alignment\_scores:

Quality: 699.50 Length: 279  
Ratio: 3.331 Gaps: 3  
Percent Similarity: 75.269 Percent Identity: 50.179

## alignment\_block:

US-09-357-675C-22 x ABL12225

Align seg 1/1 to: ABL12225 from: 1 to: 1495

```

45 ValAlaValAlaCysGlnValThrSerThrProAsnLysGlnGluAsnPhely 61
   :::::::::: ::::: ::::: ::::: :::::
146 ATGCGTGGGTGAGATGCGCTCTACACGACGACGCGCGCTAATCTTAG 195
   :::::::::::
61 sThrCysAlaGluLeuValGlnGlnAlaAlaArgLeuGlyAlaCysLeuA 78
   :::::::::::
196 CCAAGTATAGACTAGTGGATGAGGCCCAAGTCACAGAACGCTGCATGTC 245
78 lApheLeuProGlnAlaPheAspPheIleAlaArgAsnProAlaGluThr 94
   :::::::::::
246 TCTTCTGCTGAGTGTGACTTTGTGGCGCAGAGCCGACCCCAACT 295
95 LeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyLntYrSerG1 111
   :::::::::::
296 ATTGACCTCTCCGAGGCGCTTGAGCGGCGAGTAAAGCCGACATACCGGGA 345
111 nLeuAlaArgGluCysGlyLleThrPleuSerLeuGlyGlyPheHisGluA 128
   :::::::::::
346 ATTGCGGAGTGCACACAAAATTGATTTCCCTGGGTGGCTGCACGAGC 395
128 rG1GlyLnsPTrGlnAsnGlnLysIleTyrAsnCysHisValLeu 144
   :::::::::::
396 GG.....AACGATCAAAAATCTTCAACGCTCATGTTTG 430
145 LeuAsnSerLysGlySerValValAlaSerTyrArgLysThrHisLeu 161
   :::::::::::
431 CTCACAGAAAGGAGGACTAGACAGCATATACAGAAAGCTGCACATGTT 480
161 sAspValGluIleProGlyGlnGlyProMetArgGluSerAsnTyrThrL 178
   :::::::::::
481 TGAATGTACACTAA...GAGGTTCGCTACGCGAATCGATACAGTTA 527
178 ysProGlyGlyThrLeuGluProValLysThrProAlaGlyLysVal 194
   :::::::::::
528 CGCGGGGATCTGCTTGGACGCCAGTGCAGTCTGAGTGGCCAGATA 577
195 GlyLeuAlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysL 211
   :::::::::::
578 GGCCTTCAGATTTGCTACGACGCTGTTTGGTGCACCGCGGTGCTGT 627
211 uAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySer 228
   :::::::::::
628 CAGGAAGCTGGGTGCCAATTTGTTAAATACATACCCATCCGATTCACATAC 677
228 alThrGlyProAlaHisTyrGlyValLeuLeuArgAlaAlaGlu 244
   :::::::::::
678 CAACCGGTAAAGGCGCACTGGGAATCTCTCGGGGCGCAGACCATAGAG 727
245 SerGlnCysTyrValIleAlaAlaGlnCysGlyArgHisHisGluTh 261
   :::::::::::
728 ACTCAATGCTTGTGGTGGCTGCGCTCAGATGAGTTGGCACAACAGAA 777
261 rArgAlaSerTyrGlnHisSerMetValValaAspProTyrGlyThrVal 278
   :::::::::::
778 GCGACAGAGTGGGCGCAGCATGATCGTTAGCCCTGGGGAACGTTAC 827
278 AlaAlaArgCysSerGlnGlyPro...GlyLeuCysLeuAlaAlaGlyLe 293
   :::::::::::
828 TGCGTACATGCACAGCAGACGAGACTGATATATAGCAGCGCGAGGTGAG 877
294 LeuHisPheLeuGlnImetaArgGlnHisLeuProValPheGlnHisAr 310
   :::::::::::
878 CTTCGCTGCTCAATCTCTGTATCAGACCATCCCTGCTTCGACATATCG 927
310 gArgProArgLeuTyrGlySerLeuGlnHisProLeu 322
   :::::::::::
928 TCGAAAGGACATCTACGCGCTTAAACGCGTAAACGCTT 964

```

seq\_name: /SIDSI/gcdata/hold-geneseq/gene-seq-n-emb1/NA2001B.DAT:ABL12224

seq\_documentation\_block:

ID ABL12224 standard; CDNA; 3548 BP.

XX ABL12224;

XX AC 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR P-PSDB; ABB68121.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX PT interactions -

XX PS Claim 1; SEQ ID NO 31154; 21pp + Sequence listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

XX CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX CC useful in developmental biology and in elucidating cell signalling and

XX CC cell-cell interactions in higher eukaryotes for the development of

XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

XX CC sequences (AB101840-AB16175) and the encoded proteins

XX CC (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WPI

XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SO Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;

alignment\_scores:

Quality: 699.50 Length: 279

Ratio: 3.331 Gaps: 3

Percent Similarity: 75.269 Percent Identity: 50.179

alignment\_block:

US-09-357-675C-22 x ABL12224

Align seg 1/1 to: ABL12224 from: 1 to: 3548

```

45 ValAlaValAlaCysGlnValThrSerThrProAsnLysGlnGluAsnPhely 61
   :::::::::::
1199 ATGCGTGGGTGAGATGCGCTCTACACGACGACGCGCGCTAATCTTAG 1248
61 sThrCysAlaGluLeuValGlnGlnAlaAlaArgLeuGlyAlaCysLeuA 78
   :::::::::::
1249 CCAAGTATAGACTAGTGGATGAGGCCCAAGTCACAGAACGCTGCATGTC 1298
78 lApheLeuProGlnAlaPheAspPheIleAlaArgAsnProAlaGluThr 94

```



```

1299 TCTTTCGCTGAGTGTGCTGCTTTGGGCGAGAGCCGACCCCAACT 1348
      |||
95 LeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyIntYrSergI 111
      :|:
1349 ATTGAGCTCTCGAGGCGCTTGACGGGAGTAAATGGCGGAGTACCGGGA 1398
      :|:
111 nLeuAlaArgIuLucysGlyIleTrpLeuSerLeuGlyGlyPheHsGua 128
      :|:
1399 ATTGGCGAGTGCACAAATTTGATTTCCCTGGGTGGCTGCACGAGC 1448
      :|:
128 rGlyGlnAspTrpGluInAsnGlnYrSleYrAsnCyHisValLeu 144
      :|:
1449 GG.....AACGATCAAAAATCTCAACGCTCACTTTTG 1483
      :|:
145 LeuAsnSerLysGlySerValValAlaSerYrArgYrGlyThrHisLeuCy 161
      :|:
1484 CTCACAGCAAGAAAGGGAACAGTACAGTAAACAGAAAGCTGCACATGTT 1533
      :|:
161 SASpValGluIleProGlyGlnGlyPrometArgGluSerAspYrTrL 178
      :|:
1534 TGATGTTACGACTAA...GAGGTTCCCTACGCGGAATCAGATACAGTTA 1580
      :|:
178 YSPProGlyGlyThrLeuGluProProValLysThrProAlaGlyVal 194
      :|:
1581 CGCGGGATACTGCTTGAGCGCCAGTACAGACTCCAGTTGGCCAGATA 1630
      :|:
195 GlyLeuAlaIleCytyrAspMetArgPheProGluLeuSerLeuYsLe 211
      :|:
1631 GGGCTCAGATTGCTAGACCTCGTTTCTGTAAGCGCGGCTGCTCT 1680
      :|:
211 uAlaGlnAlaGlyAlaGluIleLeuThrYrProSerAlaPheGlySerY 228
      :|:
1681 CAGGAACCTGGGTCCAAATTTGTTAACTACCAATCCGATTCACATACG 1730
      :|:
228 aThrGlyProAlaHisTrpGluValLeuArgAlaArgAlaIleGlu 244
      :|:
1731 CAACGGTAAGCGGCGCTGAGGAAATCCCTCGCGGCGGAGCCATAGAG 1780
      :|:
245 SerGlnCytyrValIleAlaAlaGlnGlyGlyArgHisHsGluTh 261
      :|:
1781 ACTCAATGCTTTGTGTCGCTGCGCTCAGATAGTTGGTGCACAAACAGAA 1830
      :|:
261 rArgAlaSerYrGlyHisSerMetValAlaAspProTrpIleThrVal 278
      :|:
1831 GCGACAGAGTTGGGCGCACAGATATGTTAGCCCTGGGAAACGTAC 1880
      :|:
278 aAlaArgCySerGluGlyPro...GlyLeuCySerLeuAlaArgIleAsp 293
      :|:
1881 TGGCTGACTGCAGGAGCAGAGAGCTTGATATAGCAGCGCGAGAGTGCAC 1930
      :|:
294 LeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPheGlnHisAr 310
      :|:
1931 CTTTCCGCTTCAATCTCTGTATCAGACATGCCCTGCTTGCAGAACATCG 1980
      :|:
310 gArgProAspLeuYrGlySerLeuGlnHisProLeu 322
      :|:
1981 TCGAAGACACATCTACGCTTAAACAGCTTACACCTT 2017
      :|:
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL20222
seq_documentation_block:
ID ABL20222 standard; DNA; 5692 BP.
XX
XX ABL20222;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX

```

```

OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 12139; 21bp + Sequence Listing: English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (AB157737-AB172072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;
XX
alignment_scores:
Quality: 699.50 Length: 279
Ratio: 3.331 Gaps: 3
Percent Similarity: 75.269 Percent Identity: 50.179
Percent Similarity: 75.269 Percent Identity: 50.179
alignment_block:
US-09-357-675C-22 x ABL20222 ..
Align seg 1/1 to: ABL20222 from: 1 to: 5692
45 ValAlaValCyGlnValThrSerThrProAsnYrGlnGlnAspPheL 61
:|:
2858 ATTGCTGTGGGTGATGCGCTTACACGACAAAGCGGCTTAATCTTAG 2907
:|:
61 sThrCySAlaGluLeuValGlnGlnAlaAlaArgLeuGlyAlaCyLeuA 78
:|:
2908 CCAAGTATATAGCTAGTATGATAGGCGCAAGTACACAGAAAGCTGCAATGC 2957
:|:
78 lArPheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaGluThr 94
:|:
2958 TCTTCTCTGCTGAGTGTGCTGCTTGTGGCGGAGGCGGACCCCAACT 3007
:|:
95 LeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyIntYrSergI 111
:|:
3008 ATTGAGCTCTCGAGGCTTGACGCGCAAGTAAATGGCGCACTACGGGA 3057
:|:
111 nLeuAlaArgIuLucysGlyIleTrpLeuSerLeuGlyGlyPheHsGua 128
:|:
3058 ATTGGCAAGTGCACAAATTTGATTTCCCTGGGTGGCTGCACGAGC 3107
:|:
128 rGlyGlnAspTrpGluInAsnGlnYrSleYrAsnCyHisValLeu 144
:|:
3108 GG.....AACGATCAAAAATCTTCAACGCTCACTTTTG 3142
:|:
145 LeuAsnSerLysGlySerValValAlaSerYrArgYrGlyThrHisLeuCy 161
:|:

```





Ratio: 2.384 Gaps: 7  
Percent Similarity: 60.858 Percent Identity: 38.062

alignment\_block:  
US-09-357-675C-22 x AAD25457 ..

Align seg 1/1 to: AAD25457 from: 1 to: 1292

```

30 ArGProlaqlgthlmetSerSerSerThSerrProLuleuProleuValal 46
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 CGTGGAAAGGCATGCTC.....ACCTTCGGCTGGCCCTCATACA 74
46 aValGyslnalThrSerThrProAsnLysGlnGlnAspNheYsthrC 63
   ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 GCTT...CAAGTTTCTTCATT.....AAATCAGATACCTTACCCTGG 115
63 ysAlaGluLeuValGlnGlnLualaalaargleuGlyAlaCysLeuAlaPhe 79
   CTTTGAGCCTGATGCGGGAGGACACAAAGCAAGTCGCCAACATAGATTCT 165
116 CTTTGAGCCTGATGCGGGAGGACACAAAGCAAGTCGCCAACATAGATTCT 165
80 leuProGluAlaPheAspPheIleAlaargspnroblagluThrLeuLe 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166 CTGCGCTGAGTGCCTCAAT.....TCTCCATATGGCAACACACTTA 203
96 uleu.....SerGluProLeuAsnGlyAspLeuLeuGlyGlnTys 110
   ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 CTTTCCTGACATATGACAGAGAGATTCCTCGAGAGATCCACACAAAGCTT 253
110 ergLLeuAlaIargGluCysGlyIleThrLeuSerLeuGlyGlyPheHis 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 CTGAGAGTGCACAAAGAGACAGACATATATGTCATTGGAGGCTCCATCCCT 303
127 GluAlargGluAspTrpGluAsnGlnLysIleThrAsnGlyHisVal 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 GAA.....GAGGATGCTGGGAGAACTGATATACCTGCTC 338
143 lLeuLeuAsnSerLysGlySerValAlaIaSerTyrArgLysThrlSL 160
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 TGTGTTGGGCGTGATGGAAGATTACGTGATGACACAGAGAGATCATC 388
160 euCysAspValGluIleProGlyGlnGlyProMetIargGluSerAsnTyr 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 TGTGTGACATATGATGTCCTCGGAAATATACGTTTCAAGATCTTAAAACA 438
177 ThrLysProGluGlyThrLeuGlnProProValLysThrProAlaGly 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 TTGAGACCTGGGATAGATTCTCC...ACATTGATACGCTTACAGCA 485
193 sValGlyLeuAlaIaIeCysTyrAspMetIargPheProGluLeuSerLeu 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
486 AGTGGCGCTGGGACATCTGATGATATGCGCTTGGCGAGCTTGCAAAA 535
210 yLeuAlaGlnAlaGlyIleGluIleLeuThrTyrProSerAlaPheGly 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
536 TGTATGCAACAAAGAGCTGCCACGCTCTGGGTGATACCTGGAGCTTTCAAT 585
227 SerValThrGlyProAlaHisTyrGluValLeuLeuIargAlaIaIa 243
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 CTGACACAGAGACAGCCCACTGGGAGCTGTTCAAGGACCCGGGGCTGT 635
243 eGluSerGlnCysTyrValIleAlaAlaIaGlnCysGlyArgHisHisG 260
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 TGAATATCAGGTATGTGTGGCTACAGGCTCTCTGCTCGGGATGACAAAG 685
260 luthArgAlaSerTyrLysIleSerMetValValAspProTyrGlyThr 276
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
686 CCTGTATGTGGCTGGGGACACAGCACTGTGTGGATCTTGGGGGCAAG 735
277 ValValAlaIaIaCysSerGlnGlyProGlyLeuCysLeuAlaIaIa 793

```

|||||  
786 CCTGAACAGACTGCGTGAATTCGCGAACAATCCCATTTTAAACAGA 835  
310 rgargProAspLeuTyr 315  
::||| |||||  
836 AACGACGACACCTCTAT 852

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA81479

seq\_documentation\_block:  
ID AAA81479 standard; DNA; 69936 BP.

AC AAA81479;

DT 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence gnm\_27 SEQ ID NO:27.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

KW antigen; vaccine; diagnosis; infection; antibacterial; identification;

KM Meningococcus B; Menb; ds.

XX Neisseria meningitidis.

PN WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR ) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,

PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

PI Rappuoli R, Pizzi M;

XX WPI: 2000-318079/27.

XX Claim 7; Page 547-567; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic

XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414

XX represent specifically claimed Neisseria meningitidis genomic DNA

XX sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent

XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to

XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the

XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to

XX AAA81453 represent Neisseria meningitidis Menb polynucleotide ORF

XX sequences, which are all used in the exemplification of the present

XX invention. The nucleic acid sequences, protein sequences, and antibodies

XX against them, can be used in the manufacture of a composition. The

XX composition can be used as a medicament (or in the manufacture of a

XX medicament) for treating, preventing or diagnosing infection due to

XX Neisserial bacteria. For example, some of the identified proteins could

XX be components of vaccines against Meningococcus B; against all serotypes;

XX and/or against all pathogenic Neisseriae. Identification of sequences

XX from the bacterium will also facilitate production of biological probes,

XX particularly organism-specific probes. Attempts to make efficacious

XX Meningococcus B vaccines have failed mainly due to antigen tolerance.

XX Multivalent vaccines have also been tried but none have successfully

XX overcome antigenic variability. The provision of further, complete

XX sequences may provide an opportunity to identify secreted or surface

XX exposed proteins that may be presumed targets for the immune system and

XX which are not antigenically variable or at least more conserved than

XX other more variable regions.

SO Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

alignment\_scores: quality: 410.00 length: 275

Ratio: 2.343 Gaps: 8

Percent Similarity: 63.636 Percent Identity: 37.455

alignment\_block:

US-09-357-675c-22 x AAA81479/rev ..

Align seg 1/1 to reverse of: AAA81479 from: 1 to: 69936

45 ValAlaValCysGlnValThrSerThrProAsnGlnGluAsnPhenyl 61

27128 GTTCCGCCGCTGTCAGATGCTGCGCGCTGCGCGGCAACACGTCGC 27079

61 sThrcysAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeu 78

27078 CGCCATGAACGCTGTCGCGCGCGCGCGGACAGCGTGGATTCGG 27029

78 IapLeuLeuProGluAlaPheAspPheIleAlaArgAsnProAlaGluThr 94

27028 TGTGCTGCGCCGAAATATGCGTGTGATGCGCGCAACGATACGCAAA 26979

95 LeuLeuSerGluProLeuAsnGlyAspLeuGlnGlyIleTyr.... 109

26978 CTGCGCGCTGCGCGAGCTTTGGCGCG.....GACCGTTTCACAC 26938

110 .....SerGlnLeuAlaArgGlyCysGlyIleTyrLeuSerLeuGly 124

26937 GGCATTGACGCAACGCGGAAAGATGCGCGTGTGCTG...TTGCGCG 26891

124 Iy.....PheHisGluArgGlyGlnAspTyrGluGlnAsn 135

26890 GGAAGTGTGCGCGCTGCAAAAGTCGAGCGCGCT..... 26859

136 GlnIstIleTyrAsnCysHisValLeuLeuAsnSerIstGlySerVal 152

26858 ...AAAGTATGATATACCGCTGTGTGTCGCGGAGCGGATTAAGGAC 26812

152 IAlaSerTyrArgIstThrHisLeuCysAspValGluIleProGlyGln 169

26811 GGGGCTGACACAAAGACACCTCTTGGCTTTTCCGTTTGGCGGA. 26763

169 IyProMetArgIstSerAsnTyrThrIstProGlyIleTyrLeuGluPro 185

26762 ...CGCTATGCGGAGCGATACATCCGCGCGGCGGAGATGTG...CCG 26718

186 ProValIstThrProAlaGlyValGlyLeuAlaIleCysTyrAspMe 202

26717 CACTGTGCGGAGGAGGCGTCCGCGGCGGCGGCAATTTTACGATGT 26668

202 IArgPheProGluLeuSerLeuIstLeuAlaGluAlaGluIleT 219

26667 CCGCTTCCCAATTTTCCGACCC.....CAGTGGCGTTTACGAT 26624

219 eutThrProSerAlaPheGlySerValThrGlyProAlaHisTyrGlu 235

26623 TGATGTGCGCGCGTGTGTCGACACAGCGGCGGCAAGCGCATTTGGAG 26574

236 ValLeuLeuArgAlaArgAlaIleGluSerGlnCysTyrValIleAla 252

26573 CTGCTGTCGCGCGCGCTGCGCGGCAAAACCAATTTACGTGCGCGC 26524

252 AAlaGlnCysGlyArgHisGluThrArgAlaSerTyrGlnHisSer 269

26523 GGCACAGGCGGCTTTCGACGAAACGACGCGCATTCGCGACACGA 26474

269 etValAlaAspProTyrPheValValAlaArgCysSerGluGlyPro 285

26473 TGATGTGCGATCGTGGGCGACGTTGTGACGATTCGCCAGGCGCA 26424





```

      seq_name: /SIDSI/gcsgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI3787
seq_documentation_block:
ID   ABLI3787 standard; cDNA; 898 BP.
xx                                     BLAST
AC   ABLI3787;
xx                                     BLAST
DE   26-MAR-2002 (first entry)
xx                                     BLAST
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 35843.
KM   Drosophila: developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
xx                                     BLAST
OS   Drosophila melanogaster.
XX   WO200171042-A2.
PN   27-SEP-2001.
PD   23-MAR-2001; 2001WO-US09231.
PF   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX                                     BLAST
PA   (PEKE ) PE CORP NY.
XX                                     BLAST
PI   Venter JC, Adams M, Li PWD, Myers EW;
DR   WPI: 2001-656860/75.
DR   P-PADB; ABB69684.
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX                                     BLAST
PS   Claim 1; SEQ ID NO 35843; 21bp + Sequence Listing; English.
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL1840-ABL16175) and the encoded proteins
CC   sequences (ABB57737-ABB72072).
CC   (ABB57737-ABB72072)..
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX                                     BLAST
SO   Sequence 898 BP; 243 A; 236 C; 231 G; 188 T; 0 other;

alignment_scores:
Quality:    389.50          Length:        288
Ratio:     2.094           Gaps:         6
Percent Similarity: 64.583    Percent Identity: 33.681

alignment_block:
US-09-357-675C-22 x ABLI3787 ..
Align seg 1/I  to: ABLI3787  from: 1  to: 898

36 SetSerThserTrpgtGluLeuProLeuValAlaValCysglInvalThrse 52
|||||
:::|||||

```

```

50 AACAAACTAGG... AACTATCATGCGTTTGGCGCTGCTGCAGCTCAAGG 96
52 rthrProaInlysGlnGlnaInrheLysrThrCysAlaGlnValGlnG 69
   :::: :|||
97 TTCAGAGACAAAGTGGCCATGTGTCCAAACCGCGCTCCCAAAATGAGG 146
69 InAlaAlaLarg... LeuLylAcyLeuAlaPheLeuProGlnAlaPhe 84
   |||:::
147 CGCGCGTCAAGAGCATTAACCCCGATTGTACTCTGCGCGAGTGGTTT 196
85 AsPhe..... IleAlaArgAsnProAlaLutThrLeu 96
   ::::
197 AATGTCATCAAGCGACCAACTACTTTCGGGAGTACGCGAGACAAFT.. 244
96 uLeuSerGluProLeuAsnGlyAsrPheLeuGlnTyrSerGlnLeuA 113
   :|||
245 ..... CCGGATCGCTATACCTCGCAACACCTCCCAATTTGG 281
113 IaArgGlnCysGlyLerPheLeuGlnGlyLyrPheInsglnArgGly 129
   ||||:::
282 CAGAGACAGCACCGGTGTACATCTGGGCGGACAAATCTCGTAACGTGG 331
130 GlnAspTrgGlnGlnAsnGlnLysLlEtyrAsnCysInValLeuLeuAs 146
   ::|||:::
332 ..... GAAGACGATGCATGTACACACGCTGCAGCGTTGGTC 369
146 nSerLysGlySerValAlaLaserTyrArgLysThrInLysCysAsp 163
   : :|||:::
370 ACCCACTGCTGACCTTGGCGCAAGCATCCGCAATGCATCATCTTTGACA 419
163 aGlnLlLerProGlnGlnLyrProMetLArgLysAsnTyrThrLysPro 179
   ::::
420 TGAATGTCAAGGCTGCATCTCGTTCAAGAGTCTGAA... ACGCTGTCC 466
180 GlyLylThrLeuGluProProValLysThrProAlaLysValGlyLe 196
   :|||:::
467 GAGGCAATGATTCACATCATCACTACAGTAATGGCCACAAAGTTGGCAT 516
196 uAlaLlCysTyrAspMetArgPheProGlnLeuSerLeuLysLeuAlaG 213
   :|||:::
517 CGGCATTCGCTACGATTTTCGATTCGAGAGATGGCGAGCTCTATTCGA 566
213 InAlaGlyAlaGlnLlLeuThrTyrProSerAlaPheGlySerValThr 229
   :|||:::
567 ACGCAGCGCTCGAGATGATCATCTATCCGCTGCATTCAACATGACCACT 616
230 GlyProAlaInsrTrgValLeuLeuThrGlyLarGlyAlaLlLerGlySerG 246
   ||||
617 GGTCCACTGCACCTGGGAGACTTATGCAAGAGATCCGCTGCATATGACAAACA 666
246 nCysTyrValLlLeaAlaAlaGlnCysGlyLarGlnInsrInsglnThrArg 263
   :|||:::
667 ACTCTTTGGTGCACATCATCCAGCGCGCGAGACAAAGCGCGAGATG 716
263 lAsrTyrGlnInsrMetValValAsrProTrpGlyThrValValAla 279
   :|||:::
717 TAGCCTATGGCCATTCATGTGGTGGTGAATCCATGGCGCAAGGTGCAGAC 766
280 ArgCysSerGlnGlyProGlyLeuCysLysThrArgLylLarLeuInsrPh 296
   :|||
767 ACTGCCACTGAAGGCGAGAAATGTGTGTGCGCCATTAATTTTCTCCGA 816
296 eLeuGlnGlnMetArgGlnInsrLeuProValPheGlnInsrArgLarPro 313
   :|||:::
817 GGTGGAGCGAGTGGCGTACAGAGATTCCCGTTTGGCGCAAGAGCTGAG 866
313 sPheLylGlySer 317
   |||||
867 ATCTGTACGCCACC 880

```







OM of: US-09-357-675c-22 to: EST:\* out\_format : pfs  
Date: Apr 29, 2002 10:01 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

```
-MODEL=firmer.p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09357675/rnatc_29042002_091756_3146/app.query.fasta.1.1208  
-DB=EST -OFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.000 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELCP=6.000  
-DELEX=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdt  
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09357675 -GCNT_1.6497  
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPHY -WAIT -THREADS=1
```

## Search information block:

Query: US-09-357-675c-22  
Query Length: 323  
Database: EST:\*  
Database sequences: 13736207  
Database length: -1841457050  
Search time (sec): 2545.690000

## score\_list:

Sequence	Strd	Ordt	Zscore	EScore	Len	Documentation
gb_hlc:AK004988	+	1726.00	3060.33	3.2e-161	2993	AK004988 Mus musculus adult ma
gb_hlc:AK002269	+	1705.00	3031.84	1.3e-159	1374	AK002269 Mus musculus adult ma
gb_est1:BI647584	+	1314.00	2336.77	6.5e-121	832	BI647584 603278164F1 NIH.GC.AP_K
gb_est1:BG63013	+	1302.00	2314.75	1.1e-119	869	BG63013 602828020F1 NCI.GC.AP_K
gb_est1:AL529152	+	1265.00	2248.76	5.2e-115	844	AL529152 AL529152 LTI.NFL001.NE
gb_est1:BI332807	+	1260.50	2240.43	1.5e-115	863	BI332807 602985469F1 NCI.GC.AP_K
gb_est1:BI327680	+	1224.50	2176.77	5.3e-112	801	BI327680 602978783F1 NCI.GC.AP_K
gb_est1:AL520768	+	1221.00	2166.26	1.4e-111	890	AL520768 AL520768 LTI.NFL004.NE
gb_est1:BF780562	+	1169.00	2076.98	1.9e-106	821	BF780562 602105103F1 NCI.GC.AP_K
gb_est1:BG636916	+	1151.00	2045.12	1.1e-104	793	BG636916 602488424F1 NIH.GC.18
gb_est1:BE853807	+	1140.00	2027.96	1.0e-103	638	BE853807 ux32906.Y1 Soares.thym
gb_est1:BE22844	+	1120.50	1988.60	1.6e-101	928	BE22844 603040166F1 NIH.GC.11
gb_est1:BF385518	+	1116.00	1982.44	3.6e-101	789	BF385518 602047063F1 NCI.GC.AP_K
gb_est1:BE67049	+	1104.00	1959.47	6.8e-100	893	BE67049 DRABTC11 Rat DRG.Libra
gb_est1:BI769604	+	1075.00	1911.70	5.6e-97	992	BI769604 603054985F1 NIH.GC.12
gb_est1:AL520767	+	1075.50	1907.14	5.6e-97	992	AL520767 AL520767 LTI.NFL004.NE
gb_est1:BF780847	+	1060.00	1883.35	1.3e-95	966	BF780847 602105735F1 NCI.GC.AP_K
gb_est1:BF137474	+	1060.00	1883.69	1.1e-95	666	BF137474 601780658F1 NCI.GC.AP_K
gb_est1:BE287010	+	1057.00	1878.87	2.1e-95	654	BE287010 601097244F1 NCI.GC.AP_K
gb_est1:BF780924	+	1048.50	1862.12	1.8e-94	744	BF780924 602105436F1 NCI.GC.AP_K
gb_est1:AL522373	+	1047.50	1857.33	3.3e-94	593	AL522373 AL522373 LTI.NFL004.NE
gb_est1:BF322101	+	1034.00	1838.79	3.6e-93	593	BF322101 maa34B02.Y1 NCI.GC.AP_K
gb_est1:AL559163	+	1031.00	1829.03	1.2e-92	861	AL559163 AL559163 LTI.NFL008.TC
gb_est1:BF6762506	+	1001.00	1776.65	1.0e-89	764	BF6762506 602733954F1 NIH.GC.49
gb_est1:BF135664	+	1001.00	1774.21	1.4e-89	940	BF135664 601780826F1 NCI.GC.AP_K
gb_est1:BI477226	+	999.00	1773.96	1.5e-89	708	BI477226 602960758F1 NCI.GC.AP_K
gb_est1:BI711300	+	980.00	1766.25	1.4e-89	625	BI711300 1d96f12.X1 Human.insu
gb_est1:BG703176	+	980.00	1738.61	1.4e-87	790	BG703176 602680822F1 NIH.GC.95
gb_est1:BF703176	+	954.00	1694.73	3.8e-85	626	BF703176 601223106F1 NCI.GC.AP_K
gb_est1:BF371411	+	954.00	1699.99	4.7e-85	726	BF371411 601223207F1 NCI.GC.AP_K
gb_est1:AL565604	+	945.00	1674.96	4.8e-84	853	AL565604 AL565604 LTI.NFL003.NE
gb_est1:BI757823	+	944.00	1671.95	7.0e-84	946	BI757823 603030404F1 NIH.GC.11
gb_est1:BG94701	+	934.00	1659.18	3.6e-83	610	BG94701 603039070F1 NIH.GC.94
gb_est1:BE626803	+	933.00	1657.94	4.2e-83	582	BE626803 un11f07.Y2 Soares.mamm
gb_est1:AA086823	+	926.00	1646.19	1.9e-82	544	AA086823 m033067.11 Soares.mamm
gb_est1:BI340308	+	918.00	1631.48	1.3e-81	561	BI340308 365533 MARC.2PIC.Sus.s
gb_est1:BF968272	+	917.50	1622.03	2.9e-81	900	BF968272 602286548F1 NIH.GC.84
gb_est1:BI755493	+	911.50	1613.47	9.8e-81	813	BI755493 603027381F1 NIH.GC.11
gb_est1:BE629497	+	902.00	1603.94	4.3e-80	509	BE629497 un38f09.Y1 Soares.mamm
gb_est1:AL580840	-	901.00	1596.40	1.1e-79	830	AL580840 AL580840 LTI.NFL008.TC

gb\_est1:AL182190 + 900.00 1600.13 7.0e-80 519 | AL182190 uc64f01.r1 Soares.m  
gb\_est1:AA209049 + 899.00 1597.91 9.3e-80 538 | AA209049 mw73h11.r1 Soares.m  
gb\_est1:AA396793 + 894.50 1589.89 2.6e-79 536 | AA396793 vb28h09.r1 Soares.m  
gb\_est1:BI752623 + 889.00 1576.14 1.5e-78 746 | BI752623 603028477F1 NIH.GC  
gb\_est1:BE306735 + 882.00 1566.88 5.0e-78 564 | BE306735 601104244F1 NCI.GC.A

seq\_name: gb\_hlc:AK004988

## seq documentation block:

LOCUS AK004988 2993 bp mRNA linear HTC 19-JAN-2002  
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched  
library, clone:J300013F05:nitrilase 1, full insert sequence.  
ACCESSION AK004988  
VERSION AK004988.1 GI:12836590  
KEYWORDS HTC: CAP trapper  
SOURCE Mus musculus (strain: C57BL/6J) adult male liver cDNA to mRNA,  
clone: J300013F05.  
ORGANISM Mus musculus

## REFERENCE

1 (sites)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL MEDLINE 99279253  
PUBMED 10349636

## REFERENCE

2 (sites)  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL MEDLINE 20499374  
PUBMED 11042159

## REFERENCE

3 (sites)  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,  
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K.,  
Yui, Y., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,  
Yoneda, Y., Ishikawa, T., Ozawa, Y., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL MEDLINE 20530913  
PUBMED 11076861

## REFERENCE

4 (sites)  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
JOURNAL MEDLINE 11076861  
PUBMED 11076861

## REFERENCE

5 (bases 1 to 2993)  
Adachi, T., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,  
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Butt, C.,  
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,  
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Horii, F.,  
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S.,  
Kuribara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,  
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,  
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
Schimi, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yamanaka, T.,  
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and  
Hayashizaki, Y.

## REFERENCE

Direct Submission  
Submitted (10-JUL-2000) Yoshinide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

## TITLE

JOURNAL

Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,  
url:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5' GAGAGAGAGAGCGCGCCGCACTCGAGTTTCTTTTCTTTTCTTN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. Second  
strand cDNA was prepared with the primer adapter of sequence[5'  
GAGAGAGAGAGAGATTCAGATTAATTAATTAATTAACCCCCCTT 3']. cDNA was  
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI, 3' end:  
XhoI. Host: SOLR.

## FEATURES

## source

Location/Qualifiers  
1..2993  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="MGD:MGI:1895066"  
/db\_xref="taxon:10090"  
/clone="1300013F05"  
/sex="male"  
/tissue\_type="liver"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
169..1041  
/gene="Ntli"  
169..1041  
/gene="Ntli"  
/note="data source:MGD, source key:MGI:1350916,  
evidence:ISS  
nitrlase 1  
putative"  
/protein\_id="BAB23723.1"  
/db\_xref="GI:12836591"  
/translation="MSSSTSWELPLVAVCQVTSFVNPKQENFKTCAELVQBARLQACT  
ALPEAFDPIARNRPAETLLSEPLNDLGOYQSLARECIWISLGFHFRGDMDQN  
OKTYNCHVLINSGSVASTRKTHLDVETLGGPMKESNTTPGGTLEPVPZTPAK  
VGLAICYDMKPELPSLKAQGAELITYPSAFSVGPALMEVLLRLARLASECCYYIA  
AAOCGRHETRASVGHSMVYDPWGTAVARCSESPGLCLARIDHLQDMRQHLPIVFOH  
RRDLYGSLGHLPS"  
2973..2978  
/note="putative"  
2993  
/note="putative"  
polysite  
BASE COUNT 785 a 739 c 747 g 722 t  
ORIGIN

## alignment\_scores:

Quality: 1726.00 Length: 322  
Ratio: 5.377 Gaps: 0  
Percent Similarity: 99.689 Percent Identity: 99.689

## alignment\_block:

US-09-357-675c-22 x AK004988 ..

Align seg 1/1 to: AK004988 from: 1 to: 2993

2 LeuGlyPheIleThrArgProHisGlnLeuLeuCysThrGlyTyrAR 18  
|||||  
73 CTGGGCTTCATCACCAGGCTCTCCACCACTCGTGTACCGGATCCG 122  
18 gLeuLeuArgIleProValLeuCysThrGlnProArgProArgThrmts 35  
|||||  
123 ATTAAGTGAACCCAGACTTGTGTACAGCCGAGCCAGAACATGT 172  
35 erSerSerThrSetrPoliLeuProLeuValAlaValCysGlnValThr 51  
|||||

173 CCTCATCAACTTCTGGGAGCTGCCCTGGTGGCTGTGTGCCAGGTACA 222  
52 SerThrProAsnLysGlnLysAsnPheLysThrCysAlaGluLeuValGI 68  
|||||  
223 TCAACACCAACAAACAGACAGACATTTAAACATGTGCTGATTTGGTTCA 272  
68 nGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheA 85  
|||||  
273 AGAGCTGCACAGCTGGGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 322  
85 sPheIleAlaIArgAsnProAlaGluThrLeuLeuLeuSerLysProLeu 101  
|||||  
323 ACTTATATGACAGAAACCTGCCGAGACATTAATCTCTGCCAACACACAG 372  
102 AsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAlaIArgGlyCysGlyY 118  
|||||  
373 AATGGGAGATCTTTGGCCAAATATAGCCAGCTTGGCAGGAATGTGGAAAT 422  
118 eTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnA 135  
|||||  
423 CTGGCTGTCTTGGCGGGTTCACAGAGCGGTGCCAAGACTGGAGACAGA 472  
135 sngLysLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlySerVal 151  
473 ATCAGAAATCTACAAATGTCTCATGTGCTTTTGAAACAGCAAGGATCAGTA 522  
152 ValAlaSerTrpArgLysThrHisLeuCysAspValGluIleProGlyL 168  
|||||  
523 GTGGCCAGTTACAGAAAGACACATCTGTCCATGTGAGATCCAGGTCA 572  
168 nGlyProMetArgLysSerAsnTyrThrLysProGlyGlyThrLeuGluP 185  
|||||  
573 GGGCGCGATGAGAGAAAGCACTATACCAAGCTGGAGCAGCTTTGAGAC 622  
185 roProValLysThrProAlaGlyLysValGlyLeuAlaIleLysTyrAsp 201  
|||||  
623 CACCTGTCAAGACACCGGTGGCAAGTTGTGTGCAATCTGTTATATGAC 672  
202 MetArgPheProGluLeuSerLeuLysLeuAlaGlnAlaGlyAlaGluI 218  
|||||  
673 ATGCGCTTCCCGAATCTTTCTTGAATTTGGCTCAAGCTGGGGCAGAAAT 722  
218 eLeuThrTyrProSerAlaPheGlySerValThrGlyProAlaHisTrpG 235  
723 ACTTACTATCTCTCAGCCTTTGGATCTGTCTACAGTCCAGCCCACTGGG 772  
235 LuValLeuLeuArgAlaArgAlaIleGlnSerGlnCysTyrValIleAla 251  
|||||  
773 AGGTCTGTCTGGGGCCGCCCATTTGAATTCAGTGTATGTAAATAGCA 822  
252 AlaAlaGlnCysGlyArgHisHisGlnThrArgAlaSerTyrGlyHisSe 268  
|||||  
823 GCAGCGCACTGTGAGCGCCACCATGAACAAAGACAGATTATAGGCCATAG 872  
268 IweValValAspProTyrGlyThrValValAlaIArgCysSerGluGlyP 285  
873 CATGGTGTGACCCGTGGGCGACAGTGGTGGCCGCTGCTCCGAGGGAC 922  
285 roGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuGlnIleMetArg 301  
|||||  
923 CAGGCTCTGCTTCTCGAATTTGATCTCCACTTTCTCAACAGATGAGGC 972  
302 GlnHisLeuProValPheGlnHisArgArgProAspLeuTyrGlySerLe 318  
|||||  
973 CAAACACTGCTGCTGTTCAGACAGCCGACAGACTGACTATAGGACAGTCT 1022  
318 uGlyHisProLeuSer 323  
1023 GGGTCATCCACTCTCT 1038  
seq\_name: gb\_hnc:AK002269  
seq\_documentation\_block:

LOCUS	AK002269	1374 bp	mRNA	linear	HTC 19-JAN-2002			
DEFINITION	Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610007B07:nitrilase 1, full insert sequence.							
ACCESSION	AK002269							
VERSION	AK002269.1	GI:12832130						
KEYWORDS	HTC; CAP trapper.							
SOURCE	Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone:libRIKEN full-length enriched mouse cDNA library clone:0610007B07.							
ORGANISM	Mus musculus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
AUTHORS	Carninci,P. and Hayashizaki,Y.							
TITLE	1 (sites)							
JOURNAL	High-efficiency full-length cDNA cloning							
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)							
PUBMED	99279253							
REFERENCE	10349636							
AUTHORS	2 (sites)							
TITLE	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.							
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes							
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)							
PUBMED	20499374							
REFERENCE	11042159							
AUTHORS	3 (sites)							
TITLE	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujikake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.							
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer							
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)							
PUBMED	20530913							
REFERENCE	4 (sites)							
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.							
TITLE	Functional annotation of a full-length mouse cDNA collection							
JOURNAL	Nature 409, 685-690 (2001)							
MEDLINE	5 (bases 1 to 1374)							
PUBMED	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hata,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Hori,F., Hume,D., Imochani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Nunazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schirral,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.							
JOURNAL	Direct Submission							
MEDLINE	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.go.jp, URL: <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> , Tel:81-45-503-9222, Fax:81-45-503-9216)							
PUBMED	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details.							
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to							

prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGACAGACAGCGCCGCACACTCGAGCTTTTATTTTTTTTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAAGAGATCCAAAGAGCTCAATTATTAATTATAAACCCCCCCC 3']. cDNA was cleaved with XhoI and SfiI. Cloning sites, 5' end: SfiI; 3' end: XhoI. Host: SOLR.

```

FEATURES
    source             Location/Oualifiers
                        1..1374
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="WGI:181931"
                        /db_xref="taxon:10090"
                        /clone="061007B07"
                        /sex="male"
                        /tissue_type="kidney"
                        /clone_lib="RIKEN full-length enriched mouse cdna library"
                        /dev_stage="adult"
                        1..1374
                        /gene="Nrl1"
                        /gene="Nrl1"
                        1..1374
                        /note="data source:MCD, source key:WGI:1350916, evidence:ISS nitrilase 1"

BASE COUNT      330 a       359 c       342 g       343 t
ORIGIN

alignment_scores:
    Quality: 1705.00          Length: 323
    Ratio: 5.328              Gaps: 1
Percent Similarity: 99.071   Percent Identity: 99.071

alignment_block:
US-09-357-675C-22 x AK002269 ..

Align seg 1/1 to: AK002269 from: 1 to: 1374

2 LeuGlYPheLIeThrArgProPHISgInLeuLcysThrGLyYrAr 18
|||||
119 CTGGGCTTCATCACCGGCCCTCTCCACCAACTCTGTGAACCGAATACC 168
18 gReuLeuAtRGILeProValleucYsthrGlnP-roARgProARgThrmets 35
|||||
169 ATTACTTCGAACCCAGTACTTTGTGTACTAGCCACGACCAGAACCATGT 218
35 eRSeSerTherSertPrGluLeuProLeuValAlaValCySGlnValThr 51
|||||
219 CCTCATCACTTCCTCGGAGCTGCCCTCGTGCTGTGTCACGATTAACA 268
52 SerThProAsnLySGlnGlnUasnPhelySThCYAlaGluLeuValGI 68
269 TCACACCCAACCAAGACAGACAACTTTAAACAATGTCTGAGTTGGTTCA 318
68 ngLuAlaAlarGLeuGLYAlAcYSLeuAlaPhelaspProGluAlAlpha 85
|||||
319 AGAGGCTGCCAGACTGGTGCTTGCTGCGCTTCTTGCTGAGGCATTGG 368
85 sPHeHelleAlArGAsnP-roAlaGLUThrLeuLeuLeuSerGluProLeu 101
|||||
369 ACTTATTTCAGAGAAACCTGCCGAGACATTACTCTGTGCCAAGACATG 418
102 AsngLYAspleuLeuGLyGlnTYrSerGlnLeuAlaArgGLyCs GLY 118
|||||
419 AATGGGAGACTTTTGGCCAAATATACCAAGCTTCCACAGGGAAGATGTGGA 468
118 leThreusertleuglylphheisgluararglylglnasprtpglugn 134
|||||
469 TCTGGCTGTCTTGGCGGCTTTCACAGAGCGTGGCCAAAGCTGGAGACAG 518
```

135 AsnGlnIlySIeTyAsnCysHisValLeuLeuAnsSerLySgLySerVa 151  
|||||  
519 AATGGAATAATCTACAATTGTGATGCTGCTTTGAACAGACAGGATCACT 568  
|||||  
151 lValAlaSerTyArGlySthrHisLeuCyAspValGluIleProGlyG 168  
|||||  
569 AGTGCCCGATGACAGAGACACATCTGTGGATGTAGATCCAGAGTGC 618  
|||||  
168 lInglyProMetArGluSerAnsTyThrLySthrProGlyGlyThrLeuGlu 184  
|||||  
619 AGGGGCCATGAGAGAAAGCAACTATACCAAGCCGGAGGACCTTTGAG 668  
|||||  
185 ProProValLySthrProAlaGlyLySValGlyLeuAlaIleCysTyAs 201  
|||||  
669 CCACCTGTCAAGACACCGGTGGCAAGGTGTGTACGAATCTGTTATGA 718  
|||||  
201 pMetArGheProGluLeuSerLeuLySleuAlaGlnAlaGlyAlaGluI 218  
|||||  
719 CATGGGTTCCCTGAACTTCTTTGAAATGTGGCTCAAGCTGGGCGAAGA 768  
|||||  
218 lLeuThrTyProSerAlaPheGlySerValThrGlyProAlaHisTrp 234  
|||||  
769 TACTTACTTATCTTCACGCTTTAGATCTGTACAGGTCCAGCCCACTGG 818  
|||||  
235 GluValLeuLeuAlaArgAlaIleGluSerGlnCysTyTyValIleAl 251  
|||||  
819 GAGGTGCTGTGGGGCCCGCCCATTCATCTCAGTCTATGTAAATGAC 868  
|||||  
251 AlaAlaAlaGlnCysGlyArGHisHisGlyThrArgAlaSerTyGlyHis 268  
|||||  
869 AGCAGCGGAGTGTGAGCCACCATGAAACAGACAGATTATGGCCATA 918  
|||||  
268 erMetValValAspProTrpGlyThrValAlaIleArgCysSerGly 284  
|||||  
919 GCATGCTGTTGACCCGCGGACAGTGTGGCCCGCTGCTGCGAGGGA 968  
|||||  
285 ProGlyLeuCySLeuAlaArgIleAspLeuHisPheLeuGlnIleMetAr 301  
|||||  
969 CCAGGCTCTGCTGCTGCTGATGATCTCCACTTCTACACAGATGCG 1018  
|||||  
301 gGlnHisLeuProValPheGlnHisArgProAspLeuTyGlySerL 318  
|||||  
1019 CCAAACCTGCTGTGTCAGCAGCCGACGACTGACTTATGGCAGTGC 1068  
|||||  
318 euGlyHisProLeuSer 323  
|||||  
1069 TGGGTCACTCACTCTCT 1085  
|||||  
seq\_name: gb\_est2:BI647584  
seq\_documentation\_block:  
LOCUS BI647584 832 bp mRNA linear EST 12-SEP-2001  
DEFINITION 603278164F1 NIH\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5318631 5',  
mRNA sequence.  
ACCESSION BI647584  
VERSION BI647584.1 GI:15561820  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 832)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:

<http://image.lnl.gov>  
Plate: L1801807 row: 0 column: 16  
High quality sequence stop: 810.  
Location/Qualifiers  
FEATURES  
source  
1..832  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5318631"  
/clone\_1id="NIH\_CGAP\_Mam3"  
/tissue\_type="tumor, gross tissue"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; site:1: NotI;  
site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert 2 kb. Library constructed by Life  
Technologies, catalog #12017-018. Investigators providing  
samples: Lohar Hennighausen/Chu-Xia Deng, NIH Reference  
for transgenic model: Xu et al., Nature Genetics 22, 37-43  
(1999). Note: this is a NCI\_CGAP Library."  
BASE COUNT 198 a 214 c 217 g 203 t  
ORIGIN  
alignment\_scores:  
Quality: 1314.00 Length: 263  
Ratio: 5.073 Gaps: 1  
Percent similarity: 98.479 Percent identity: 98.479  
alignment\_block:  
US-09-357-675C-22 x BI647584 ..  
Align seg 1/1 to: BI647584 from: 1 to: 832  
1 MetLeuGlyPheIleThrArProBrOnHisGlnLeuLeuCySthrGlyTy 17  
|||||  
48 ATGCGGGGCTTCATCACCAGGCGCTCCTCACCACAACTCTGTACCGGATA 97  
|||||  
17 rArgLeuLeuArGlyIleProValLeuCySthrGlnProArGProArGThm 34  
|||||  
98 CCGATTACTTCCAAATCCAGTACTTGTACTCAGCCACAGCCGACAGAACCA 147  
|||||  
34 eTserSerSerThrSerTrpGluLeuProLeuValAlaValCysGlnVal 50  
|||||  
148 TGTCTCTCATCACTTCTGTGGAGCTGCCCTGTGTGCTGTGGCCAGGTA 197  
|||||  
51 ThrSerThrProAsnLySglnGluAsnPhelySthrCysAlaGluLeuVa 67  
|||||  
198 ACATCAACACCAACACAGACAGAACTTTAAACATGTGCTGAGTGGT 247  
|||||  
67 lGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaP 84  
|||||  
248 TCAGAGGCTGCCAGACTGGGTGCTGCTGCTGCTTGTGCTGAGGCAT 297  
|||||  
84 heAspPheIleAlaArgAsnProAlaGluThrLeuLeuLeuSerGluPro 100  
|||||  
298 TTGACTTATTTCACAGAAACCTGCCGAGACATTACTCTGTCGCAACA 347  
|||||  
101 LeuAsnGlyAspLeuLeuGlyGlnTySerGlnLeuAlaArgGluCysG 117  
|||||  
348 CTGAATGGGATCTTTGGCCAAATATACCAAGCTTGCAG. GAATGTGG 396  
|||||  
117 ylleTrpleuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluG 134  
|||||  
397 AATCTGGCTGTCTTGGCGGTTTCCACAGAGGTGGCCAAAGACTGGGAC 446  
|||||  
134 lAsnGlnIlySIeTyAsnCysHisValLeuLeuAnsSerLySgLySer 150  
|||||  
447 ACAAATCAGAAATCTACAAATGTGATGCTTTTGAACGCAAGGATCA 496  
|||||  
151 ValValAlaSerTyArGlySthrHisLeuCyAspValGluIleProG 167  
|||||  
497 GTAGTGGCCAGTTACAGAGAAATACATCTGTGAGTGTAGATCCACAG 546  
|||||  
167 yGlnGlyProMetArGluSerAnsTyThrLySthrProGlyGlyThrLeuG 184

```
|||||
547 TCAGGGCCCATGAGAGAAAGCAACTATACCAAGCCTGAGGCACTCTTG 596
184 lnpProvallysthrProalaglylyValglyLeuAlaIleCysTyr 200
|||||
597 AGCCACTGTCACAGACACCGCTGGCAAGGTGGTCAGCAATCTGTAT 646
201 AspMetArpPheProgluLeuSerLeuLyLeuAlaIleAlaIleValaI 217
|||||
647 GACATGCGGTCCCTGAACTTCTTGAATTGGCTACGCT.GGGGCAGA 695
217 uileuThrTyrProSerAlaPheGlySerValThrGlyProAlaHisT 234
|||||
696 AATACCTACTATCTTACCTCAGCTTTGGATCTGTACAGTCCAGCCACT 745
234 rpluValleuLeu.ArgAlaArgAlaIleGluSerGlnCysTyrValI 250
|||||
746 GGGAGGTGCTGCTGGCGGCCCGCCCATTTGAATCTCAGTCTATGTAA 795
250 eAlaAlaIleGlnCysGlyArgHisGluThrArg 262
|||||
796 AGCAGCAGCCCACTGTGGACGCCACATTTGAACACAGG 832
seq_name: gb_est2:BG963013

seq_documentation_block:
LOCUS BG963013 869 bp mRNA linear EST 12-JUN-2001
DEFINITION 602828020F1 NCI_CGAP_Co24 Mus musculus cDNA IMAGE:4982813 5',
ACCESSION BG963013
VERSION BG963013 GI:14350650
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 869)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10986 row: 9 column: 06
High quality sequence stop: 805.
Location/Qualifiers
1. 869
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4982813"
/lab_host="NCI_CGAP_Co24"
/clone_lib="DHI0B (T1 phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 Kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 212 a 221 c 233 g 202 t 1 others
ORIGIN
alignment_scores:
Quality: 1302 00 Length: 276
Ratio: 4.932 Gaps: 2
Percent Similarity: 95.652 Percent Identity: 94.203
alignment_block:
```

```
US-09-357-675C-22 x BG963013
Align seg 1/1 to: BG963013 from: 1 to: 869
1 MetLeuGlyPheIleThrArgProHisGlnLeuLeuCysThrGlyTyr 17
|||||
47 ATGCTGGGCTTCATCACAGAGCCTCTCCACCACTCTGTGTACCGGATA 96
17 rArgLeuLeuArgIleProValLeuCysThrGlnProArgProArgTyrM 34
|||||
97 CCGATTACTTGTGAATCCAGTACT.TGTACTACACCCAGGCCAGAAACA 145
34 etSerSerSerThrSerTrpIleuProLeuValAlaValCysGlnVal 50
|||||
146 TGTCTCATCACTTCTCTGGAGAGCTGCCCTGTGTGCTGTGTGCAGSTA 195
51 ThrSerThrProAsnLysGlnGluAsnPhelSerThrCysAlaGluLeu 67
|||||
196 ACATCAACACCAACAAGCAAGAGAACTTTAAACATGTGCTGAGTTGT 245
67 lGlnGluAlaIleAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaP 84
|||||
246 TCAAGAGGCTGCGACACTGGGTGCTTGCCTGCTTCTTGCCTAGAGCAT 295
84 heAspPheIleAlaArgAsnProAlaGluThrLeuLeuSerGluPro 100
|||||
296 TTGACTTTATTGCAAGAAACCTGCGGACATTAATCTCTGTCCGACCA 345
346 CTGAATGGGAGATCTTNTGGCCCAATATAGCCAGCTTGCAGGGAATGTGG 395
101 LeuAsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAlaArgGluCysG 117
|||||
117 yIleTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGlu 134
396 AATGTGGCTGTCTTGGCGGCTTTCACAGACGTGGCCAAACCTGGGAGC 445
134 lnaGlnGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlySer 150
|||||
446 AGAATCAGAAATCTACATTTGTATGTCTTTTGAACAGCAAGGAGATCA 495
151 ValValAlaSerTyrArgLysThrHisLeuCysAspValGluIleProG 167
|||||
496 GTAGTGGCCACTTACAGAAACACATCTGTGCATGTAGAGATCCAGG 545
167 yGlnGlyProMetArgGluSerAsnTyrThrLysProGlyGlyThrLeuG 184
546 TCAGGGGCCAATGAGAAACCACTATACCAAGCTTGAGGCACTCTTG 595
184 lnpProvallysthrProalaglylyValglyLeuAlaIleCysTyr 200
|||||
596 AGCCACTGTCAGAGACACCGGCTGGCAGAGTTGGTCTAGCAATCTGTAT 645
201 AspMetArpPheProgluLeuSerLeuLyLeuAlaIleAlaIleValaI 217
|||||
646 GACATGCGGTCCCTGAACTTCTTGAATTGGCTCAAGCTGGGGCAGA 695
217 uileuThrTyrProSerAlaPheGlySerValThrGlyProAlaHisT 234
|||||
696 AATACCTACTATCTTACCTCAGCTTTGGATCTGTACAGTCCAGCCACT 745
234 rpluValleuLeu.ArgAlaArgAlaIleGluSerGlnCysTyrValI 250
|||||
746 GGGAGGTGCTGCTGGCGGCCCGCCCATTTGAATCTGACGCTATGTAT 795
250 eAlaAlaIleGlnCysGlyArgHisGluThrArgAlaSerTyrGln 267
|||||
796 AGCAGCAGCGCAGTGTGAGCCCA.CATGAACAAGACCAAGTATATGCC 844
267 is.SerMetValValAspProTyr 274
|||||
845 ATAGACATGTGTGTGAACCGTGG 868
seq_name: gb_est1:AL529152
```

```

seq_documentation_block:
LOCUS      AL529152               844 bp      mRNA      linear      EST 13-FEB-2001
DEFINITION AL529152 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSDD004YM06 5
ACCESSION  AL529152
VERSION    AL529152.1  GI:12792645
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 844)
AUTHORS   Li, W. B., Gruber, C., Jesse, J., and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.
FEATURES
     source
         1..844
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CSDD004YM06"
            /clone_1bp="LTI_NFL001_NBC4"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /note="Organ: brain. Vector: PCMVSPORT 6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the PCMVSPORT 6
            vector. Library was normalized. Library was constructed
            by Life Technologies. Contact: Feng Liang Life
            Technologies, a division of Invitrogen 9800 Medical Center
            Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
            8371 Email : filiang@litech.com URL :
            http://fulllength.invitrogen.com"

BASE COUNT      182 a      238 c      235 g      189 t

alignment_scores:
    Quality: 1265.00      Length: 263
    Ratio: 5.000          Gaps: 0
    Percent Similarity: 96.198      Percent Identity: 87.833

alignment_block:
US-09-357-675c-22 x AL529152 ..

Align seg 1/1 to: AL529152 from: 1 to: 844

```

```

144 euLeuAsnSerIysGlySerValAlaIaSerTyrArgIysThrHisLeu 160
|||||.....
251 TCCTGAACGCAAGGAGGAGTAGTGCCACTTACAGGAAGACACTCTG 300
161 CysAspValGluIleProGlyInGlyProMetArgGluSerAsnTyrTh 177
|||||.....
301 TGTGACGTAGACATTCACAGGCGGCTATGTGTGGAAGCAACTCTAC 350
177 rLysProGlyGlyThrLeuGluProProValIysThrProAlaGlyLys 194
|||||.....
351 CATGCCGTGGCCAGACTCTGATGACCTGTACAGCACACAGGAGGAGCA 400
194 aIGlyLeuAlaIleCysTyrAspMetArgPheProGlyLeuSerLeuLys 210
|||||.....
401 TGTGCTTACTGTCTGCTATGACATCGGGTCCCTGAACTCTCTTGCA 450
211 LeuAlaGluAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySe 227
|||||.....
451 TTGGCTCAAGCTGGAGCAGACAGATCTTACCTTACCTTCAGCTTTGGATC 500
227 rValThrGlyProAlaHisTyrProIleValLeuLeuArgAlaArgAlaIleG 244
|||||.....
501 CATTCAGGCCAGCCAGCCACTGGAGGAGTGTGCTGGCGGCCGCTATGCG 550
244 IuSerGlnCysTyrValIleAlaAlaAlaGlnCysGlyTyrGlnHisGlu 260
|||||.....
551 AAACCCAGTCTATGTAGTGAGCAGCAGCAGTGTGAGCCACACCTAGAG 600
261 ThrArgAlaSerTyrGlyHisSerMetValAlaAspProTyrProIleTyrA 277
|||||.....
601 AAGAGAGCAAGTTATGACCACACAGATGCTGTAGACCCCTGGGAGACAT 650
277 lValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspL 294
|||||.....
651 GGGGCCCTGCTGTGAGGAGGAGCCCTGCTGCTGCTGCTGCTGCTGCTG 700
294 euHisPheLeuGlnGlnMetArgGlnHisLeuProAlaPheGlnHisArg 310
|||||.....
701 TCACACTATGTGGACAGCTGGCCGACACCTGCTGTGTCACGACCCG 750
311 ArgProAspLeuTyrGlySerLeuGlyHisProLeuSer 323
|||||.....
751 AGGCTGACCTCTATGCAATCTGGGTCAACCCACTCTCT 789

seq_name: gb_est2:B1332807

seq_documentation_block:
LOCUS      B1332807               863 bp      mRNA      linear      EST 30-JUL-2001
DEFINITION 602985469P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5138032 5',
ACCESSION  B1332807
VERSION    B1332807.1  GI:15017464
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 863)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cga@p5-remai.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: L14M1339 row: j column: 17
            High quality sequence start: 22
            High quality sequence stop: 825.

```





```

58 ATGCTGGGCTTCAATCACAGGCTCTCTACCAACTCCTGTGTACCGAGTA 107
17 TATGLeuLeuAArgIleProValLeuCysThrGlnProArgProArgThrM 34
108 CGATATTACTTGAATCCAGTACTTGTACTGACGCCAGGCCAGACCA 157
34 eSerSerSerThrSerThrPgluLeuProLeuValAlaValCysGlnVal 50
158 TGTCCATCACTCACTTCTGTGAGCTGCCCTGTGCTGTGCTGACGTA 207
51 ThrSerThrProAsnLysGlnGlnAsnPhelLysThrCysAlaGluLeu 67
208 AATTCACACCAACACAGAGAACTTTAAACATGTGCTGAGTTGGT 257
67 TGlncLualaAArgLeuGlyAlaCysLeuAlaPhelLeuProGluAlaP 84
258 TCAAGAGGCTGCCAGACTGGTGTCTGCTGCGCTTGTGCTGAGAGCAT 307
84 hAspPheIleAlaArgAsnProAlaGluThrLeuLeuSerGluPro 100
308 TTGACTTATTTGCACGAAACCTGCGCAGACATTAATCTGTCGGAACA 357
101 LeuAsnGlyAspLeuLeuGlnGlnThrSerGlnLeuAlaArgGluCysG 117
358 CTGAATGGGGATCTTTGGGCCAATATAGCCAGCTTGCAG, GAATGTGG 406
117 YLleTrrPLeuSerLeuGlyGlyPheHISgluaArgGlyInAspTrrpGlu 134
407 AATTCGGCTGTCTTGGGGGTTTCCACAGAGCTGGCCAAAGACTGGAGAG 456
134 InAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlySer 150
457 ACAATCAGAAATCTCAATTTGATGATGCTTTTGAACAGCAGAGGATCA 506
151 ValValAlaSerTyrArgLysThrHisLeuCysAspValGluLeuProG 167
507 GTAGTGGCCAGTTACGAGAAACATCTGTGATGTAGAGATCCAGG 556
167 YGlnGlyProMetArgGluSerAsnTyrThrLysProGlyGlyThrLeuG 184
557 TCAGGGGGCCAAATGAGAGAAAGCACTATACCAAGCTGGAGGCACTTGG 606
184 LrpProValLysThrProAlaGlyLysValGlyLeuAlaIleCysTyr 200
607 ACCCACTCTCAAGACACCGCTGCGCAAGTGTGTACCAATCTGTAT 656
201 AspMetArgPheProGluLeuSerLeuLysLeuAlaGlnAlaGlyLagi 217
657 GACATGGGCTTCCCTGAATCTTCTTGAATGGCTCAAGCTGGGGCAGA 706
217 uLleLeuThrTyrProSerAlaPheGlySerValThrGlyProAlaHisT 234
707 AATACTTACTTATCTTCAAGCTT, GAATCTGTACAGGTCCAGGCCACT 755
234 TpgLuvAlleuLeuAArgAla, IleGluSerGlnCysTyrVal 249
756 GCGAGGTGCTGCTGGGGGCC...GGCCATTGAATCATCATGTATGTA 800
seq_name: gb_est1:AL520768
seq_documentation_block:
LOCUS AL520768 890 bp mRNA linear EST 13-FEB-2001
DEFINITION AL520768 LTI_NFL004_NBC2 Homo sapiens cDNA clone CSDB002YD05 5
ACCESSION prime, mRNA sequence.
VERSION AL520768.1 GI:12784261
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 890)
AUTHORS Li, W. B., Gruber, C., Jesse, J. and Polayes, D.

```

```

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..890
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDB002YD05"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 191 a 251 c 241 g 206 t 1 others
ORIGIN
alignment_scores:
Quality: 1221.00 Length: 277
Ratio: 4.770 Gaps: 2
Percent similarity: 92.419 Percent identity: 83.032
alignment_block:
US-09-357-675c-22 x AL520768 ..
Align seg 1/1 to: AL520768 from: 1 to: 890
1 MetLeuGlyPheIleThrArgProProHisGln.....LeuLeuCy 14
||||||| ||||||| ||||||| ||||||| |||||||
60 ATGCTGGGCGCTCATTCACAGGCTCTCTCACAGATTCTGCTTCTGTG 109
14 sThrGlyTyrArgLeuLeuArgIleProValLeuCysThrGlnProArgp 31
110 TCTTGAGCTCCGATACCTCACTCACTTGTGCTGCTGCTGCTGCTGCT 159
31 roArgThrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 46
160 CCAGAGCCATGGCTATCTCTCTCTCTCCGCGAAGTGGCCCTGTGGCT 209
47 ValCysGlnValThrSerThrProAsnLysGlnGlnAsnPhelThrCy 63
210 GTGTGCCAGGTACATCGAGCCAGACACCAAGCAAGACTTTAAACATG 259
63 salAgluLeuValGlnLualaAArgLeuGlyAlaCysLeuAlaPhel 80
260 TGTCTGAGCTGTTTGAAGAGCTGCGAGACTGGGTGCTGCTGCTTTC 309
80 euProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 96
310 TGGCTAGGCAATTGACTTATGCAAGGACCTTCACAGAGAGCTAACAC 359
97 LeuSerGluProLeuAsnGlyAspLeuGlyGlnTyrSerGlnLeuAl 113
360 CTGTCTGAACCACTGGGTGGGAAACTTTTGAAGAATATACCACTTGC 409
113 aArgGluCysGlyIleTrrPLeuSerLeuGlyGlyPheHISgluaArgGly 130
410 CAGGGAATGTGGACTGTGCTTGTGGGTGTTTCAATGACGAGGCC 459
130 InAspTrrpGluInAsnGlnLysIleTyrAsnCysHisValLeuAsn 146
||||||| ||||||| ||||||| ||||||| |||||||

```

460 AAGACTGGAGCAGACTCAGAAAATCTACAATGTGTCGTCGTGAAC 509  
147 SerLysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspVal 163  
510 ACCAAAGGGGAGTACGTGGCCACTTACAGAGAACACATCTGTGTGACGT 559  
163 TgLuLeuProGlyGlnGlyProMetArgLysSerAsnTyrThrLysProG 180  
560 AGAGATTCCAGGAGGAGGCGCTATGTGTGAAGCAACTCTACCATGCTG 609  
180 TgLyThrLeuGlnProProValLysThrProAlaGlyLysValGlyLeu 196  
610 GGCCAGCTCTTGAGTACCTGTGACACACACAGCAGCAAGATTGGTCTA 659  
197 AlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysLeuAla 213  
660 GCGTGTGTATGACATGGGTTCCCTGAACTCTCTGCGCATGGCTCA 709  
213 naIaGlyAlaGluIleuThrTyrProSerAlaPheGlySerValThrG 230  
710 AGCTGAGCAGAGATACCTTACCTTCTTCAGCTTTGGATCCATTACAG 759  
230 TgProAlaHisTgPgluValLeuLeuArgAlaArgAlaIleGluSerGln 246  
760 GCCCAGCCCACTGGAGGTGTCTGCGGGCCGCTGCTATCGAACCCAG 809  
247 CysTyrValIleAlaAlaAlaGlnCysGlyArgHisGluThrArgAl 263  
810 TGTATGTGTGTGGCAGCAGCAGTGTGGACGCCCATGAGAAAGAGAGC 859  
263 aserTyrGlyHisSerMetValValAspPro 273  
860 AAGTTATGGCCACAGCATGTGGKAGACCC 890

seq\_name: gb\_est2.Bf780562

seq\_documentation\_block: 821 bp mRNA linear EST 12-JAN-2001  
LOCUS Bf780562  
DEFINITION 602103103P1 NCI\_CGAP\_Kid14 Mus musculus cDNA IMAGE:4221289  
5', mRNA sequence.  
ACCESSION Bf780562  
VERSION Bf780562.1 GI:12085595  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 821)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LHAM9806 row: e column: 02  
High quality sequence stop: 713.  
Location/Qualifiers  
1. 821  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4221289"  
/clone\_lib="NCI\_CGAP\_Kid14"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: Kidney; Vector: pCMV-SPORT6; site\_1: NotI;  
site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Constructed by Life

BASE COUNT 196 a 213 c 213 g 199 t  
ORIGIN  
alignment\_scores:  
Quality: 1169.00 Length: 253  
Ratio: 4.733 Gaps: 4  
Percent Similarity: 97.628 Percent Identity: 96.838  
alignment\_block:  
US-09-357-675c-22 x Bf780562  
Align seg 1/1 to: Bf780562 from: 1 to: 821  
1 MetLeuGlyPheIleThrArgProProHisGlnLeuLeuCysThrGlyTyr 17  
12 ATGCTGGGCTTCATTCACAGAGCCCTCCACCAACTCCTGTGTACCGGATA 61  
17 TArgLeuLeuArgIleProValLeuCysThrGlnProArgProArgThrm 34  
62 CCGATTACTTGAATCCCACTACT .GTACTCAGCCCGAGGCCAGAACCA 110  
34 etSerSerSerThrSerTgPgluLeuProLeuValAlaValCysGlnVal 50  
111 TGTCTCATCACTTCTCTGGAGCTGCCCCGTGTGTGTGTGTGTGTGTGT 160  
51 ThrSerThrProAsnLysGlnGlnAsnPhelThrCysAlaGluLeuVal 67  
161 ACATCAACACCAAAACAGAGAACTTTAAACATGCTGCTGAGTGTGT 210  
67 TgGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAla 84  
211 TCAAGAGGCTGCCAGACTGGTCTTGGCTTCTGCTTCTGAGGAT 260  
84 heAspPheIleAlaArgAsnProAlaGlnThrLeuLeuSerGluPro 100  
261 TTGACTTATTATTCACGAAACCTGCGAGACATTACTCTGTCCGAACCA 310  
101 LeuAsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAlaArgGluCysG 117  
311 CTGAATGGGATCT .TTGGGCCAATATAGCACCTTCCAG .GAATGTGG 358  
117 TyleTgPgluSerLeuGlyGlyPheHisGluArgGlyGlnAspTgPglu 134  
359 AATCTGCTGCTCTTGGCGGTTTCCAGAGCGGCGCAAGACTGGAGAGC 408  
134 TnsngLnlYsIleTyrAsnCysHisValLeuLeuAsnSerLysGlySer 150  
409 AGAATCAGAAATCTACAAATTCATGTCTTTTGAACACAGAGGATCA 458  
151 ValValAlaSerTyrArgLysThrHisLeuCysAspValGluLeuProG 167  
459 GTAGTGCCAGGTACAGAGACACATCTGCGATGTAGAGATCCAGAG 508  
167 TgGlnGlyProMetArgLysSerAsnTyrThrLysProGlyGlyThrLeu 184  
509 TCAAGGGCCACTGAGAGAAAGCAATATACCAAGCCGGAGGCACTTGG 558  
184 TgProProValLysThrPro .AlaGlyLysValGlyLeuAlaIleCys .T 200  
559 AGCCACCTGTCAAGACACCGCTGCTGGCAAGTGTGTGTACATCTGTGG 608  
200 TgAspMetArgPheProGluLeuSerLeuLysLeuAlaGlnAlaGlyVal 216  
609 ATGACAGCGGTCCTCGAATCTTCTTGAATGGGTCAAGCGGGAGCA 658  
217 Gln .IleLeuThrTyrProSerAlaPheGlySerValThrGlyProAla 233  
659 GAAATATCTTACTTATCTTACGCTTGTGATCTGTATACAGGTCACGCC 708  
233 TgTgPgluValLeuLeuArgAlaArgAlaIleGluSer .GlnCysTyrVal 249

seq\_name: gb\_est2:BG436916

seq\_documentation\_block: 793 bp mRNA linear EST 14-MAR-2001

LOCUS BG436916

DEFINITION 602488424F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4620892 5', mRNA sequence.

ACCESSION BG436916

VERSION BG436916.1 GI:13343422

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 793)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: DCTD/DTP/Gazdar  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>  
Plate: LNCM1382 row: 0 column: 05  
High quality sequence stop: 767.

FEATURES

Location/Qualifiers

source 1..793

organism="Homo sapiens"

db\_xref="taxon:9606"

clone="IMAGE:4620892"

/clone\_lib="NIH\_MGC\_18"

/tissue\_type="large cell carcinoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pGB7; site\_1: XhoI; site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 181 a 210 c 230 g 172 t

ORIGIN

alignment\_scores: Quality: 1151.00 Length: 267  
Ratio: 4.622 Gaps: 5  
Percent Similarity: 93.258 Percent Identity: 83.895

alignment\_block: US-09-357-675c-22 x BG436916 ..

Align seg 1/1 to: BG436916 from: 1 to: 793

42 LeuprPLeuValAlaValcysGlnValThrSerThrProAsnLysGlnI 58  
|||||  
3 CTGCCCTGGCTTCTCCCTGAGGCAATTTGACTTCACTTACAGGACCT 52  
|||||  
58 uasnPhelYstnrcysAlaGluValcysGlnValThrSerThrProAsnLysGlnI 75  
|||||  
53 GAACCTTTAAACATGCTGACGTGGTTCGAGAGGCTGCCACAGCTGGG 102  
|||||  
75 lacysLeuAlaPheLeuProGlnAlaPheAspPheIleAlaArgAsnPro 91

|||||

103 CCTGCTGGCTTCTCCCTGAGGCAATTTGACTTCACTTACAGGACCT 152

92 AlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyI 108  
|||||

153 GCAGAGACGCTACCTGCTGTGAACCACTGGGTGGGAACCTTTGAGACA 202  
|||||

108 nTyrSerGlnLeuAlaArgGlycysGlyIleThrPheSerLeuGlyGlyP 125  
:|||||

203 ATACACCCAGCTTGGCCAGGAATGTGACTGTGGCTGCTTGGGGCTT 252  
:|||||

125 hehISGLuArgGlyGlnAspTrpGluGlnAsnGlnLysIleTyrAsnCy 141  
|||||

253 TCCATGAGCGGTGGCCAGGACTGGAGCACACACAGAAATTCACAAATTT 302  
|||||

142 HisValLeuLeuAsnSerLysGlySerValAlaSerTyrArgLysTh 158  
|||||

303 CAGGTGCTGTGAACGACAAAGGGGAGTAGTGCCACTTACAGAGAGAC 352  
|||||

158 rHisLeuCyaspValGluIleProGlyGlnGlyPromeTarGlySera 175  
|||||

353 ACATCTGTGTGACGTAGAGATTCCAGGGGAGGCTATGTGTGAAGACA 402  
|||||

175 snTyrThrLysProGlyGlyThrLeuGluProProValLysThrProAla 191  
|| |||

403 ACTCTACCACTGCTGGGCCAGCTTGTGACTCACTGTACAGACACAGCA 452  
|||||

192 HisLysValGlyLeuAlaIleCysTyrAspMetArgPheProGluLeu 208  
|||||

453 GCAGAGATTGGTCTGACTGTGCTATGACATGCGGTTCTCTGAACCTTC 502  
|||||

208 rLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro.Serala 224  
||| |||||

503 TCTGGATTGGCTCAAGCTGAGAGACAGATTACTTACTTACTTACAGCT 552  
|||||

225 PheGlySerValThrGlyProAlaHisTrpGluValLeuLeuArgAla A 241  
|||||

553 TTTGGATTCATTACAGAGCCACCCAGCTGGAGGTGTTCTCGGGCCCC 602  
|||||

241 rAlaIleLeuSerGlnCysTyrValIleAlaAlaAlaGlnCysGlyArg 257  
|||||

603 GGTCTATCCAAACCCAGTCTATGATGAGGACACACAGTGTGACGCC 652  
|||||

258 HSHSLGluThrArgAlaSerTyrGlnHisSerMetValValAsp.ProT 274  
|||||

653 CACCATGACAGAGAGCAAGTTATGGCCACAGCATGTGTGATACCCCT 702  
|||||

274 rPGLy.ThrValValAlaArgCysSer.GluGlyProGlyLeuCysLeuA 290  
|||||

703 GGGGAACAGAGTGTGCCCGCTCTTTGAGGGGCGAGGCTGTGCTTGC 752  
|||||

290 laArgLysPheLeuHisPheLeuGlnGlnMetArgGlnHis 303  
|||||

753 CCCGANTAGACCTCAAGTATCTGCGACAGTGTGCCGGAAC 793  
|||||

seq\_name: gb\_est2:BE853807

seq\_documentation\_block: 638 bp mRNA linear EST 27-SEP-2000

LOCUS BE853807

DEFINITION ux23906.y1 Soares\_thymus\_2NBM7 Mus musculus cDNA clone IMAGE:3472378 5' similar to YR:088526 O88526 NITRILASE HOMOLOG 1. [1] /, mRNA sequence.

ACCESSION BE853807

VERSION BE853807.1 GI:10312377

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 638)

AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index



Ratio: 4.293 Gaps: 6  
Percent similarity: 88.475 Percent identity: 78.305

Alignment block:

US-09-357-675C-22 x B1822844 ..

Align seg 1/1 to: B1822844 from: 1 to: 928

```

1 MetLeuGlyPheIleThrArgProPheHisGln.....LeuLeuGly 14
|||||.....|
39 ATGCTGGGCTTCATCATCCAGGCGCTCTCTCAGACATTCCTCTCCCTGTG 88
14 sThrGlyTyrArgLeuLeuArgIleProValLeuGlyThrGlnProArg 31
| ||| .....|
89 TCCTGGACTCCCGATACCTCAACTCTCACTTGTCTGCTCAGCCAGCG 138
31 rArGThrMetSer...SerSerThrSerThrGlnLeuProLeuValAla 46
|||||.....|
139 CCAGAGCCATGGCTATCTCTCTCTCTCTCTGCGAAGTCCCTGGTGGCT 188
47 ValCysGlnValThrSerThrProAsnGlnGlnAsnPheLeuThrCys 63
|||||.....|
189 GGTGCGCAGGTAACTACATCGACGCGACAGCAAGCAACTTTAAACATG 238
63 sAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
|||||.....|
239 TCCTGAGCGGTTCAGAGGCTGCGAGACTGGGCTGCTCTCTGCTTCC 288
80 euProGlnAlaPheAspPheIleAlaArgAspProAlaGluThrLeuLeu 96
|||||.....|
289 TCCTGAGGCAATTGACTTCATTCATGACGGGACCTGCGACAGACGCTAC 338
97 LeuSerGluProLeuAsnGlnLysAspLeuLeuGlyGlnTyrSerGlnLeu 113
|||||.....|
339 CTGTCTGAACCACTGGTGGTGGAACTTTGGAAGATACACCGAGTTCG 388
113 aArgGluCysGlyIleTrrPLeuSerLeuGlyLysPheHisGluArgGly 130
|||||.....|
389 CAGGAAATGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 438
130 LAspTrpGluGlnAsnGlnLysIleTyrAsnGlnHisValLeuLeuAsn 146
|||||.....|
439 AAGACTGGGAGCAGACTCAGAAATCTACAAATGTCACCTGCTGCTGAA 488
147 SerLysGlySerValValAlaSerTyrArgLysThrHisLeuCysAsp 163
|||||.....|
489 ACCAAGGGGCGAGTAGTGGCCACTTACAGAAAGACACATCTGTACCT 538
163 LgluIleProGlnGlyGlnGlyProMetArgGluSerAsnTyrThrLys 180
|||||.....|
539 AGAGATTCCAGGCGAG..GGGCTTATGTGTGAAGCAACTCTACCATGCC 587
180 LylGlyThrLeuGlnLupProProValLysThrProAlaGlyLysValGly 196
|| .....|
588 GCGCCAGTCTTGAAGTACCTGTCAGGCCACGACGAGCAAGATGGTCTA 637
197 AlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysLeuAla 213
|||||.....|
638 GCTGTCTGTATGACATGGGCTTCCCTGAACCTCTCTGCGAATGGCTCA 687
213 nAlaGlyAlaGluIleLeuThr..TyrProSerAlaPheGlySer...Val 228
|||||.....|
688 AGCTGGAGAGAGATACCTTACCTTACCTTACCTTACCTTACCTTACCT 737
229 ThrGlyProAlaHisThrGluValLeuLeuArgAlaArgAlaIleGly 245
|||||.....|
738 CAGGCGCCAGCCCACTGGAGGTGTGGCGGCGCCGCTATGCAAAAC 787
245 rGlnCysTyrVal..IleAlaAlaAlaGlnCysGly..ArgHisHisGly 261
|||||.....|
788 CCAGTGCATATGTTAGTGGCGCCAGCAGCACTGTGGAAGCCACCATGAGA 837
261 rArgAlaSerTyrGlnLysSerMetValValAspProTrpGlyThrVal 278

```

```

|||||.....|
838 GAGAGC..ACTTATGGGACACAGCATGTGCTAAACCTGGGCGACAGGGG 886
278 aAlaArgCysSerGlnGlyProGlyLeu 287
::|.....|
887 CCCGTCTCTGAGGGCGAGCCCTGGCTTG 915

```

seq\_name: gb\_est2:BF385518

seq\_documentation\_block:

LOCUS BF385518 789 bp mRNA linear EST 27-NOV-2000  
DEFINITION 602047063F1 NCL\_CGAP\_L19 Mus musculus cDNA IMAGE:4196559 5',  
mRNA sequence.

ACCESSION BF385518  
VERSION BF385518.1 GI:11366823

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 789)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps@email.nih.gov](mailto:cgaps@email.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
plate: LRAM533 row: n column: 16  
High quality sequence stop: 754.

#### FEATURES

Location/Qualifiers

source 1..789  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4196559"  
/clone\_lib="NCL\_CGAP\_L19"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: Liver; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCL\_CGAP Library."  
BASE COUNT 182 a 206 c 208 g 193 t  
ORIGIN

#### alignment\_scores:

Quality	Length	Gaps
Ratio: 1116.00	239	4
Percent Similarity: 96.653	Percent Identity: 92.469	

alignment\_block:

US-09-357-675C-22 x BF385518 ..

Align seg 1/1 to: BF385518 from: 1 to: 789

```

1 MetLeuGlyPheIleThrArgProPheHisGlnLeuLeuGlyThrGlyTyr 17
|||||.....|
83 ATGCTGGGCTTCATCATCCAGGCGCTCTCTCAGACATTCCTCTCCCTGTG 132
17 rArgLeuLeuArgIleProValLeuGlyThrGlnProArgProAlaGly 34
|||||.....|
133 CCGATTACTTCGAATCCAGTACT..TGTACTCAGGCCAGGCGCCAGAACCA 181
34 eTserSerSerThrSerThrGlnLupProLeuValAlaValCysGlnVal 50
|||||.....|
182 TCCTCTCATCAACTTCGAGAGCTGCCCTGTGCTGTGTCAGAGTA 231

```

```

51  ThrSerThrProAsnLysGlnGluAsnPhelYsThrCysAlaGluLeuVa
|||||
232 ACATCAACACCAACAAGCAAGAACTTTAAAAACATGCTGAGTTGCT 281
67  1GlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaP 84
|||||
282 TCAGAAGAGCTGCAGACTGGGTGCTTCTGCTGCTTCTGCTGAGGCAT 331
84  heAspPheIleAlaArgAsnProAlaGluThrLeuLeuLeuSerGluPro 100
|||||
332 TTGACTTTATTTGCGAAGAACCTGCGAGACATTAATCTCTCGGAACCA 381
101 LeuAsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAlaArgGluCysG 117
|||||
382 CTGATGGGATCTTTGGGCGCAATATAGCCAGCTTCCAGGGAATGTGG 431
117 YLeTyrPLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTyrP 134
|||||
432 AATCTGCTGTCTTGGCGGTTTCCAGAGCGTGGCCAGACTGGGAGC 481
134 1AsnGlnLys1LeTyrAsnCysHisValLeuLeuAsnSerLysGlySer 150
|||||
482 AGATACAGAAATCTACAAATGTCTATGTCTTTGAACACAGCAGGATCA 531
151 ValValAlaSerTyrArgLysThrHisLeuCysAspValGluLeuProG 167
|||||
532 GTAGTGGCCAGTTACAGAAAGACACATCTGTGCATATAGATCCACAG 581
167 YGlnGlyProMetArgLysSerAsnTyrThrLysProGlyGlyThrLeu 184
|||||
582 TCAGGGGCCAATGAGAGAAAGCAACATACCAAGCTTGAGGAGCCTCTG 631
184 1uProProValLysThrProAlaGlyLysValGlyLeuAlaIleCysTyr 200
|||||
632 AGCCACTGTCAGAACACCGCTGGCAAGGTGCTTACAGATCTGTAT 681
201 AspMetArgPheProGluLeuSerLeuLysLeuAlaGluAlaGlyLag 217
|||||
682 GACATGCCGCTTCCCTAACCTTTCTTGAAATGTGTCAGAGCTGGGG...C 728
217 1u1LeuThrTyrProSerAlaPheGlySerValThrGlyProAlaH 233
:::|||||
729 AGATATATACTATCTTCACGCTTG...GATCTGTTCAGTTACGCCA 775
233 sTrrGluValLeu 237
|||||
776 CTGGAGTGTCTGTG 788

seq_name: gb_est2:BG667049

seq_documentation_block:
LOCUS BG667049 893 bp mRNA linear EST 30-APR-2001
DEFINITION DRABTCl1 Rat DRG Library Rattus norvegicus cDNA clone DRABTCl1 5',
mRNA sequence.
ACCESSION BG667049
VERSION BG667049.1 GI:13888971
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 893)
Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,
Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and
Zhang,X.
Distinct gene expression profiles of rat dorsal root ganglion
induced by peripheral nerve axotomy
Unpublished (2001)
JOURNAL Contact: Zhang Xu
COMMENT Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China

```

```

Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
Pudong New Area, P.R.China. Please contact with Zhang Xu
(xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES
source 1. 893
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRABTCl1"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
BASE COUNT 218 a 213 c 262 g 196 t 4 others
ORIGIN

alignment_scores:
Quality: 1104.00 Length: 272
Ratio: 4.658 Gaps: 4
Percent Similarity: 87.132 Percent Identity: 82.721

alignment_block:
US-09-357-675c-22 x BG667049 ..
Align seg 1/1 to: BG667049 from: 1 to: 893

60 PheLysThrCysAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCys 76
|||||
1 TTTAAACGTGTGCTGAGCTGTTCAGAGGCTACCAAGCTGGGCGCTTG 50
76 sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsnProAlaG 93
|||||
51 CTTGGCCTTTTGCTGCTGAGCATTTGACTTATTTATGACAGAAATCTGCG 100
93 1uThrLeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyGlnTyr 109
|||||
101 AGACATTACTCTCTGCCAACCACTGATGGGAGCTTTTGGACAATAT 150
110 SerGlnLeuAlaArgGluCysGlyYLeTyrPLeuSerLeuGlyGlyPhe 126
|||||
151 AGCCAGCTTGGCCAGCAATGTGGAATCTGGCTGTCTGGGTGTTTCCA 200
126 sGluArgGlyGlnAspTyrPLeuGlnAsnGlnLysIleTyrAsnCysHisV 143
|||||
201 CGAGCGTGGCCAGACTGGAGAGACTCAGACAAATCTACAAATGTCTGTG 250
143 AlLeuLeuAsnSerLysGlySerValValAlaSerTyrArgLysThrHis 159
|||||
251 TGCCTTCGACAGCAAGGATCAGTAGGCGCAGTTACAGAGACACAT 300
160 LeuCysAspValGluLeuProGluGlnGlyPheMetArgLysSerAsnTyr 176
|||||
301 CTATGTGATGTAGAGATCCAGGTGAGGCGCTATGAGAGAAAGCAACTA 350
176 rThrLysProGlyGlyThrLeuGluProProValLysThrProAlaGlyL 193
|||||
351 TACCATGCTGTGATAGCGCTTTGAACCACTGTCAAGACACAGAGCA 400
193 yValGlyLeuAlaIleCysTyrAspMetArgPheProGluLeuSerLeu 209
|||||
401 AGTTGGTCTAGCAATCTGTATGACATGCGCTTCCCGAAGCTTCTCTG 450
210 LysLeuAlaGlnAlaGlyAlaGluLeuThrTyrProSerAlaPheG 226

```

```
|||||
451 AATATGGCTCAAGCTGGGAGAAATCTTACTTACCTTCAGCCTTGG 500
226 yserValThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAla 243
501 ATCTGTACCGGGCCAGCCCACTGGGAGTGTCTGGCGGCCGCCCA 550
243 lgluSerGlnCysTyrValIleAlaAlaGlnCysGlyArgHisHis 259
551 TTGAGTCTCAGTCTACGTATGACAGAGACAGTGTGGACCCACAT 600
260 GluThrArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 276
601 AGACAGACAGACAGATTACGGGCACACAGATGTGGTTGA..CCCTGGGAC 649
276 rValValAlaArgCysSerGlnGlyProGlyLeu..... 287
650 ATGGGGGACAGCTGTGC..GAGGGACACAGCTTGGCTGCTCAATGATCC 696
288 .....CysLeuAlaArgIleAspLeuHisPheLeu 297
697 ACTTCTGAAGAATGGCCGAATGGCGGTCAAGATCGGACATA.....CTA 740
298 GlnGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAsp 314
741 TCGAGTCGGCGCTCAAGTTACTCGGGGTGGAGACACCGC.....CT 781
314 uTyrGlySerLeuGly 319
782 CTAGGCTCGGCTTGGG 797

seq_name: gb_est2:B1769604
seq_documentation_block:
LOCUS B1769604 846 bp mRNA linear EST_25-SEP-2001
DEFINITION 60305498SF1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5',
mRNA sequence.
ACCESSION B1769604
VERSION B1769604.1 GI:15761182
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 846)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apds@email.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov
Plate: L14M11512 row: j column: 22
High quality sequence stop: 838.
Location/Qualifiers
1..846
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5204469"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SportS;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymus pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb. Insert size
range 1-3 kb. Library is normalized and enriched for
```

```
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT 179 a 240 c 226 g 201 t
ORIGIN
alignment_scores:
Quality: 1077.00 Length: 264
Ratio: 4.414 Gaps: 4
Percent Similarity: 92.424 Percent Identity: 82.576
alignment_block:
US-09-357-675c-22 x B1769604 ..
Align seg 1/1 to: B1769604 from: 1 to: 846
1 MetLeuGlyPheIleThrArgProPheHisGln.....LeuLeu 14
59 ATGCTGGGCTTCATACACAGGCTCCCTCACAGATTCCCTGCTCCCTG 108
14 sThrGlyTyrArgLeuValArgIleProValLeuCysThrGlnProArg 31
109 TCCTGGACTCCGGATACCTCACTCACTCACTTCTTGTCTCAGCCAGCC 158
31 rArgThrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 46
159 CCAGACCAATGGGTATCTCTCTCTCTCTCTCGCAACTGCCCCGTTGG 208
47 ValCysGlnValThrSerThrProAsnLysGlnGlnAsnPheLysThr 63
209 GTGTGCCAGTAACATCAGACGACGACGACCAACACAACTTTAAACATG 258
63 sAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
259 TGTGAGCTGTGTTGAGAGGCTGCCAGACTGGGTGCTCTGCTTTC 308
80 euProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeu 96
309 TGGCTGAGCATTTGACTTCATTCATTCAGCGGACCTGCAGAGCGTAC 358
97 LeuSerGluProLeuAsnGlnLysPheLeuGlyGlnTyrSerGlnLeu 113
359 CTTCTGGAACCACTGGGTGGAACTTTTGAGAAATACCACTTCGCT 408
113 argGluCysGlyIleTrpLeuSerLeuGlyGlnPheHisGluArgGly 130
409 CAGGAAATGTGACGTGCTGCTCTTGGGTGGTTTCCATGACGTGGCC 458
130 lAspTrpGluGlnAsnGlnLysIleTyrAsnCysHisValLeuAsn 146
459 AAGACTGGGAGCAGACTCAGAAATCTACAAATGTACAGCTGCTGAG 508
147 SerLysGlySerValValAlaSerTyrArgLysThrHisLeuCysAsp 163
509 ACCAAAGGGGAGTAGTGGCCACTTACAGAAACACATCTGTGTAGCT 558
163 lGluLeuProGlyGlnGlyProMetArgLysSerAsnTyrThrLysPro 180
559 AGAGATTCACAGGCGAGGGGCTATGTGTAAAGCAACTTACCATCCT 608
180 lValThrLeuGluProProValLysThrProAlaGlyLysValGlyLeu 196
609 GGCCAGCTTGTGAGTCACCTGTACAGACACACAGAGCAAGATTGGTCA 658
197 AlaIle.CysTyrAspMetArgPheProGluLeuSerLeuLysAlaG 213
659 GCTGTCACTTATGACATCGGTTCCCTCAACCTCTCTGCGCATTTGG 708
213 lValGlyValGluIleLeuThrTyrProSerAlaPheGlySerValThr 229
709 AAGCTGACACAGATTAATTACTTACCTTACGCTTTTGGATTCATTA 758
```



```
230 G1yProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluSerG1 246
||||| ||||||||||||||| |||||||||||
759 GCCCAGC.CACTGGAGGTGTCTGCGGC.CGTGCTATCGAAACCA 806
||||| |||||||||||
246 nCySTyRVaIlleAlaAla.AlaGInCySgIyArGHIs 258
||||| |||||||||||
807 GTGCTATGTAGTGGCAGCAGCACAGTGTGGAGCCAC 844
||||| |||||||||||
```

**THIS PAGE BLANK (USPTO)**

OM of: US-09-357-675c-22 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Apr 29, 2002 10:03 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL-framer.p2n.model -DEV-xlh  
-O/cgn2\_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-25 + 210.00 389.83 5.0e-14 1785  
-DB-Issued\_Patents\_NA -OPMT-fasta -SUFFIX-p2n.rml  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-LOOPEXT=0.000 -GAPEXT=4.500 -OGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX-biosum6 -TRANS-human40.cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE-pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT-pfs -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER-US09357675 @CGN1.118 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-357-675c-22  
Query length: 323  
Database: Issued\_Patents\_NA\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 85.140000

## Score list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-25 +					210.00	389.83 5.0e-14 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-63 +					210.00	389.83 5.0e-14 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-25 +					210.00	389.83 5.0e-14 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-63 +					210.00	389.83 5.0e-14 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-5 +					208.00	385.75 8.5e-14 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-5 +					208.00	385.75 8.5e-14 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-61 +					207.00	383.70 1.1e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-61 +					207.00	383.70 1.1e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-7 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-7 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-9 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-9 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-11 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-11 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-21 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-21 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-27 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-27 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-29 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-29 +					206.00	381.66 1.4e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-65 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-65 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-13 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-13 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-23 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-23 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-27 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-27 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-29 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-29 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +					205.00	379.61 1.9e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-39 +					203.00	377.24 2.5e-13 1559
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-39 +					203.00	377.24 2.5e-13 1559

seq\_name: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-25  
seq\_documentation\_block:  
Sequence 25, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Namba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 340078/1992  
 FILING DATE: 21-DEC-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Wegner, Harold C.  
 REGISTRATION NUMBER: 25,258  
 REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 672-5300  
 TELEFAX: (202) 672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 25:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1785 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 ORIGINAL SOURCE:  
 STRAIN: JM109 PAD445  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 233..1144  
 US-08-294-871A-25

alignment\_scores:  
 Quality: 210.00 Length: 338  
 Ratio: 1.304 Gaps: 14  
 Percent Similarity: 47.633 Percent Identity: 24.556

alignment\_block:  
 US-09-357-675c-22 x US-08-294-871A-25 ..

Align seg 1/1 to: US-08-294-871A-25 from: 1 to: 1785

```

46 AlavAlCysGlnValThr.....SerThrProAs 55
   |||||:|||||:|||||:|||||:|||||:|||||:
164 GCGATGTGTTCACACACGTTTCCGCGCGCTGGCGGACATCACCTAG 213
   |||||:|||||:|||||:|||||:|||||:|||||:
55 nLysGlnGluAspPheLysThr..... 62
   |||||:|||||:|||||:|||||:|||||:|||||:
214 GAGGAGCAGAGGTTCATGACACGTACATGATCTTCGACGTGGACAAAC 263
   |||||:|||||:|||||:|||||:|||||:|||||:
63 .....Cys 63
264 AAGTCGATCGCGCGCGGAGACACGCGAACAGTGTCTGCTGCTT 313
64 AlagLLeuValGlnGlnAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80
   |||||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATGCTGACGAAGCCGCGGAGCGGCGGAAATTTCATTGCTCT 363
80 uProGlu.....AlaPheAspPheLeAla 89
   |||||:|||||:|||||:|||||:|||||:|||||:
364 CCGGCAATCGCGCTTACGACCTTTCCCGCGCTGGCATTTTCACCGAGC 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
   |||||:|||||:|||||:|||||:|||||:|||||:
414 AGCGGAGCTGATAGCTTCTATGAGACCGAA...ATGCGCGCGCGG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGlyCysGlyIleTyrPheLeuSer 122
   |||||:|||||:|||||:|||||:|||||:|||||:
461 GTCCTGCTACCTTTTGAGAAAGCCCGGAAATCGGATCGCTCAATCT 510
122 u.....GlyLysPheHisGluArgGlyGln 131
   |||||:|||||:|||||:|||||:|||||:|||||:
511 GCGCTACGCTGACTGCTGCTGGAAGCGGCTCAAGCGCTCC..... 553
131 sPTpGluGlnAsnGlnLysIleTyrAsnGlyValLeuLeuAsnSer 147
   |||||:|||||:|||||:|||||:|||||:|||||:
554 .....TTCAACACGCTCATTTTGGTGATAG 580
148 LysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspValGln 164
   |||||:|||||:|||||:|||||:|||||:|||||:
  
```

```

581 TCAGCGACAGATCGTCGGCAAGTATCGTAAGATCCATTTC..... 619
164 uLleProGlyGln.....GlyProMetArg.....G 173
   |||||:|||||:|||||:|||||:|||||:|||||:
620 ...CCGGCTCACAGAGATAGAGAGCCCTACCGCGCTTCACAGATCTTG 665
173 LuSerAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
   |||||:|||||:|||||:|||||:|||||:|||||:
666 AAAAGCGTATTTCAGAGCGGCGGATCTCGCGCTCCGCTATGACAGC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPhePro 206
   |||||:|||||:|||||:|||||:|||||:|||||:
716 GACGCGCGAATAAGGATGTTCACTGCAACAGATCGCGCTGGCTGA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyValGluIleLeuThr.....T 221
   |||||:|||||:|||||:|||||:|||||:|||||:
766 AGCTTCGCGCGGTGATGGCTTCAGGGCGCGGATCATCTCGCGCGCT 815
221 TyrProSerAlaPheGlySerValThrGlyProAlaHisTyrPgluValLeu 237
   |||||:|||||:|||||:|||||:|||||:|||||:
816 ACAACACGCGGACACCAATCCCGCTTCGCCAGCAGCAGCAGCTGAGC 865
238 .....LeuArgAlaArgAlaIleGluSerGlnC 247
866 TCCTTCGACCATCTCTATGATGAGCGCGGCTTATCAACAGCGGCG 915
247 sTyrValIleAlaAlaGlnCysGlyArgHisHisGluThrArgAla 264
   |||||:|||||:|||||:|||||:|||||:|||||:
916 CTGGTCGCGCGCGCGCGGCAAGTGGGCAATG...GAGGAAATCGATTC 962
264 eTyrGlyHisSerMetValAlaAspProTyrPgluThrValAlaArg 280
   |||||:|||||:|||||:|||||:|||||:|||||:
963 TCCTGCGGCACTCTGATCGATGTCGCGCGGCGGAAATCTCGCTTC 1012
281 CysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLeuHis 296
   |||||:|||||:|||||:|||||:|||||:|||||:
1013 ACTACGACGCTGAAGACAGAGTATCACCGCGCGGCTGATCGATG 1062
296 eLeuGlnIleMetArgGlnHisLeuProValPhe...GlnHisArgArg 312
   |||||:|||||:|||||:|||||:|||||:|||||:
1063 CTGCGGGAATCGCTGAGACACATCTTCACTTCAAGCAGCATCGTCAAC 1112
312 roAspLeuTyrGly 316
   |||||:|||||:|||||:|||||:|||||:|||||:
1113 CCGACGACTATGCT 1126
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-63
seq_documentation_block:
: Sequence 63, Application US/08294871A
: Patent No. 5824522
: GENERAL INFORMATION:
: APPLICANT: Ikenaka, Yasuhiro
: APPLICANT: Nanda, Hirokazu
: APPLICANT: Takano, Masayuki
: APPLICANT: Yajima, Kazuyoshi
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
: APPLICANT: Yamada, Kazuhiko
: APPLICANT: Hiraiishi, Yoshiro
: TITLE OF INVENTION: Immobilized Enzyme Preparation and
: TITLE OF INVENTION: Process for Producing D-Amino Acid
: NUMBER OF SEQUENCES: 70
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
  
```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294.871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIORITY APPLICATION DATA:
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JMI09 PAD469
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-63

```

```

alignment_scores:
    Quality: 210.00      Length: 343
    Ratio: 1.304        Gaps: 14
    Percent Similarity: 46.939    Percent Identity: 23.907

```

```

alignment_block:
US-09-357-675C-22 x US-08-294-871A-63

```

```

Align seg 1/1 to: US-08-294-871A-63 from: 1 to: 1785
46 AlaValCysGlnValThr.....SerThrProAs 55
||||:||||| :|||
164 GCGATGTGTTTCACACACGTTTCCCGGCCGCTGGCGGACATCACCTAG 213
55 nlyGlnGluAsnPhelysthr..... 62
||||:||||:||||| |||
214 GAAGAGACAGAGTTTCATGACAGTCAGATGATCTTGACATGGGACAC 263
63 .....Cys 63
264 AAGCTCGATCGCGCGCGGAGACAGCGAAGAGTGTGCTGCTGCTT 313
64 AlaGluLeuValGlnGluAlaIaIaArgLeuGlyAlaCysLeuAlaPheLe 80
||||:||||| :|||
314 CTCGACATGCTGACGAAAGCCGAGCGCGGCCGCAATTCATGTCCTT 363
80 uProGlu.....AlaPheAspHeIleAla 89
|||||
364 CCGCGAACTCGCGCTTACGACCTTCCCGCGCTGGCATTTACACCGACG 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
||||:||||| :|||
414 AGCGCGAGCTGATAGCTTATGAGACCGAA...ATGCGCGCGCGGTG 460
106 LeuGlnGlnIlySerGlnLeuAlaArgGlyCysGlyIleThrPheSerLe 122
|||
461 GTCCGTCACACTCTTGAGAAAGCCGCGAAGCTCGGATCGGCTCAACT 510
122 u.....GlyGlyPheHisGluArgGlyAla 131
|
511 GGGCTAGCTGTAAGCTGCTGCGAAGCGCGCTGAAGCTGC..... 553
131 sPTPrGluGlnAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 147
|||||
554 .....TTCAACAGCTGCATTTGGTGGATAG 580
148 LysGlySerValValAlaIaSerTyrArgLysThrHisLeuCysAspValG 164
||||:||||:||||| |||
581 TCAGGCAAGATCGTGGCAAGTATCGTAAGATTCATTGG..... 619
164 uIleProGlyGln.....GlyProMetArg.....G 173
|||||
620 ....CCGGGTACAAAGAGTACGAGCGCTACCGCGCTTCAGCATCTTG 665
173 IuSerAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
||||:||||| :|||
666 AAAAGCGTATTTCAGCGCGGCGCATCTCGGCTCCGCTATGACGCTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
||||:||||:||||| |||
716 GACCCCGGAAATGGGATGTCATCGCAACGATCGCCCGTGGCCGCA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 223
|
766 AGCTGGCGGTGATGGCTCAGCGCGCGGAGATCATGTC..... 808
223 eAlaPheGlySerValThrGlyProAlaHisTyrGluValLeu..... 237
||||:||||| :|||
809 .....GGCGGCTACACAGCGCGGACCCACATCCCATTTGTTCCCGAG 850
238 .....LeuArgAlaIaArgAl 242
851 CACGACCACTGACGCTCTTCACACATCTCTATGATGACGAGCGGCGTC 900
242 aIleGluSerIleCysTyrValIleAlaIaIaGlnCysLysArgHisH 259
||||:||||| :|||
901 TTATCAGAACGGGGCTGTGTCGCGCGCGGCAAGGTGGGCAATG...G 947
259 IsGluThrArgAlaSerTyrGlyHisSerMetValValAspProTyrP 275
||||:||||| :|||
948 AGGAGAACTGATGCTGCTGCGCCATCTTCATCGTCCGCCGACCGCGG 997

```

```

276 тгВаVаIValdAлgуssеrвU...glPгoоlуauсуslеudAlar 291
      ::::|lllll      ::::|lll      |ll
998 GAAATGTCGCGCTCACTACGAGCCCTGGAGAAGAGGTGATCACGCCGCGC 1047
291 gtlеasPlеnIsPhеuеngInglmеtargGlnlIsleuProValphe. 307
      ::||lllll      ::::::|ll::|ll::|ll
1048 CСRGAТТССТGATGCSTGCGCGGAAATGCGTGAAACAATCTTCAACTCA 1097
308 ..GlnlIsAtgarGrpOalspleutylgly 316
      ||lllllll:::lll::|lllll
1098 AGCAGCATCTGTAAGCCCCAGCATCATGTGT 1126

seq_name: /сgn2_6/Ptdodata/2/lна/6A_COMB.seq:US-08-876-398A-25

seq_documentation_block:
Sequence 25, Application US/08876398A
Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENUKA, Yasuhiro
APPLICANT: NANBA, Hirokazu
APPLICANT: TAKANO, Masayuki
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBYLASE IMPROVED IN
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: A35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 pAD445
```

```

1  ; FEATURE:
2  ; NAME/KEY: CDS
3  ; LOCATION: join(23...1141)
4  ; US-08-876-398A-25
5
6  alignment_scores:
7      Quality: 210..00      Length: 338
8      Ratio: 1.304          Gaps: 14
9      Percent Similarity: 47.653      Percent Identity: 24.556
10
11 alignment_block:
12 US-09-357-675C-22 x US-08-876-398A-25 ..
13
14 Align seg 1/1 to: US-08-876-398A-25 from: 1 to: 1785
15
16 46 AlaValAlCysGlnValThr.....SerThrProAs 55
17  |||||..... |||||
18 164 GCGAGTGTCTTTCACACAGCTTTTCCCGCGCTGGCGCGGACATCACCTAG 213
19  |||||..... |||||
20 55 nLysGlnLysAsnPhelyThr..... 62
21  :|||:..... |||
22 214 GAAGAGAGAGAGAGGTTTCATGACACGCTGACATGATCTTGCAGTGGGACAC 263
23  :|||:..... |||
24 63 .....Cys 63
25
26 264 AAGTCCCATCGCGCGCGCGAGACAGCGACAGAGTCTGTTCTGCTT 313
27
28 64 AlaGlnLeuValGlnGlnAlaAlaArgLeuGlnValAcylSerLeuAlaPhe 80
29  |||||..... |||||
30 314 CTCGCAATGCTGACGAAAGCCGACAGCCGGCGCGCAATTTCATTGCTT 363
31
32 80 uProGlu.....AlaPheAspPheIleAla 89
33  |||||..... |||
34 364 CCCCAGATCGCGGCTTACGACCTTCTCCGCGGCTGCATTTCACCGCAG 413
35
36 89 rGAsnProAlaGluThrIleuLeuLeuSerGluProLeuAsnGlyAspLeu 105
37  :|||:..... |||||
38 414 AGCGCGAGCTCGATAGCTCTPATGAGACGAA...ATGCCGCGCGGCTG 460
39
40 106 LeuGlyGlnTySerGlnLeuAlaArgLysGlnCysGlyIleThrLeuSerIe 122
41  :|||:..... |||||
42 461 GTCCTCCACACTTTTGAGAAAGCCGCGAGACTCGGATCGGCTTCATCT 510
43
44 122 u.....GlyIlePheHisGluArgGlyLeu 131
45  |..... |||||
46 511 GGGCTACCTCAACTCGTCGTGGAAGCGCGGCTCAACCGCTGC..... 553
47
48 131 sPTPrGluGlnAsnGlnLysIleTyIAsnGlyHisValLeuLeuAsnSer 147
49  .....TTCAACACGTCATTGTTGGTGAGTAAg 580
50
51 554 .....TTCAACACGTCATTGTTGGTGAGTAAg 580
52
53 148 LysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspValG 164
54  :|||:..... |||||
55 581 TCAGCGCAAGATCGTCGCGCAAGTATCGTAAATGCATTG..... 619
56
57 164 uIleProGlyLys.....GlyProMetArg.....G 173
58  |||||..... |||||
59 620 ...CCGGGTTCACAAGAGTACGAGGCTACCGCGCTCCACCATCTTG 665
60
61 173 LuSerAsnTyrThrIlyAspProGlyLysThrLeuGlnGluProProValLysThr 189
62  :|||:..... |||
63 666 AAAAGCGTTATTTTCGAGCGCGGCGCATCTCGGCTTCCCGGCTATGACGTC 715
64
65 190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
66  :|||:..... |||||
67 716 GACGCGCGCAAAATGGGGATGTTCTATCTGCAGACATCGCGCGCTGCA 765
68
69 206 uLeuSerLeuLysLeuAlaGlnAlaGlyAlaGlnIleLeuThr.....T 221
70  :|||:..... |||||
71 766 AGCTGCGCGGTGATGGGCTTCAGGCGCGCGGAGATCATTCGCGCGGCT 815
72
73 221 yTPProSerAlaPheGlySerValThrGlyProAlaHisTPTPrGluValLeu 237

```

```

186 ACAACAGCGCGACCAATCCACCTTCCCGACGACGACCACTACG 865
238 .....LeuArgAlaArgAlaIleGluSerGlnC 247
866 TCCTTCACCATCTCTATGATGACGAGCGGCTTATGACAGCGGAGC 915
247 sTyRValIleAlaIleAlaGlnCysGlyArgHisGluThrArgAla 264
916 CTGGTCGGCGCGCGGCAAGTGGGCATG...GAGGAGAACGTCATGC 962
264 eTyRgIHisSerMetValIValAspProTrrpGlyThrValIValArg 280
963 TGCTCGGCCACTCTCTGATCTGTGGCGCGGACCGGGAATCTGCTCTC 1012
281 CysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLeuHis 296
1013 ACTACGACGCTGGAAGACGAGGTGATCACCAGCGCGCTGCATCTCATCG 1062
296 eLeuGlnGlnMetArgGlnHisLeuProValPhe...GlnHisArgArg 312
1063 CTGCGGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATGTCAGC 1112
312 roAspLeuTyRgLy 316
1113 CCCAGCACTATGCT 1126

```

seq\_name: /cgn2.6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-63

seq\_documentation\_block:

Sequence 63, Application US/08876398A

Patent No. 6083752

GENERAL INFORMATION:

APPLICANT: IKENAKA, Yasuhiro

APPLICANT: NAKABA, Hirokazu

APPLICANT: TAKANO, Masayuki

APPLICANT: YAJIMA, Kazuyoshi

APPLICANT: YAMADA, Yukio

APPLICANT: TAKAHASHI, Satomi

TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY & LARDNER

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/876,398A

FILING DATE: 16-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/211,641

FILING DATE: 11-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/JP93/01101

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 340078/1992

FILING DATE: 21-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 212692/1992

FILING DATE: 10-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wegner, Harold C.

REGISTRATION NUMBER: 25,258

```

REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JMI09 PAD469
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-63

```

alignment\_scores: Quality: 210.00 Length: 343

Percent Similarity: 46.939 Ratio: 1.304 Gaps: 14

Percent Identity: 23.907

alignment\_block:

US-09-357-675C-22 x US-08-876-398A-63 ..

Align seg 1/1 to: US-08-876-398A-63 from: 1 to: 1785

```

46 AAlaValCysGlnValThr.....SerThrProAs 55
164 GCGATGTTTTCACACGTTTCCCGCGCTGGCGCGACATCACCAG 213
55 nLySGlnGluAspNheYthr..... 62
214 GAAGAGCAGAGATTTCATGACGTCAGATGATCTTCACTGGAGCAC 263
63 .....Cys 63
264 AAGTCGATCGCGCGCGGAGACGAGAACGAGTCGTCGTTCTT 313
64 AAlaGluValGlnGlnAlaIleArgLeuGlyAlaCysLeuAlaPhe 80
314 CTCGACATGTCGACAAAGCCGCGAGCGGCGGGAATTCATTGCTT 363
80 uProGlu.....AlaPheAspPheIleAla 89
364 CCCGAACTCGCGCTTACGACCTTCTCCGCGCTGGCATTTACACGAG 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
414 AGCGGACGCTGATAGTTTATGAGACGAA...ATCGCGCGCGGCTG 460
106 LeuGlyGlnTySerGlnLeuAlaArgGlyCysGlyIleTrrpLeuSer 122
461 GTCCGTCACCTTTTGAAGAGCGCGGAACTCGGAGCTTCAACT 510
122 u.....GlyGlyPheHisGluArgGlyAla 131
511 GGGCTACGCTAAGCTCGTCGAGAGCGCGCTCAAGCGTGC..... 553
554 .....TTCAACAGCTCATTTGGTGGATAG 580
148 LysGlySerValAlaIleAspTyrArgGlyThrHisLeuCysAspValG 164
581 TCAGCAAGATCTCGCGCAAGTATCGTAAGTCAATTGG..... 619
164 uLleProGlyGln.....GlyProMetArg.....G 173
620 ....CCGGTCAACAAGAGTACGAGGCTTACCGCGCTTCCAGCATCTTG 665

```

```

173 Inserta nTyrThrLysProGlyThrLeuGluProValLysThr 189
      ||:||||| :|||
666 AAAAGCATATTTCAGACGGGCGATCGGCTCCCGGCTATGACGTC 715
190 ProAlaGlyValGlyLeuAlaIleCysTyrAspMetArgPheProG1 206
      |||:||||| :|||
716 GACGGCGGCAAAATGGGATTCATCTGCACAGATCGCGCGGCTGA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyLeuIleLeuThrTyrProS 223
      | :|||
766 AGCCTGGCGGCTATGCGCTCAGGGCGCGGATCATCTGC..... 808
223 e rAlaPheGlySerValThrGlyProAlaHisTrpGluValLeu..... 237
      |||:| :|||
809 .....GGCGGCTACACACCGCCAGCCACATCCCATGTTCCCGCAG 850
238 .....LeuArgAlaArgAl 242
851 CACGACCACTGACGTCCTCCACCATCTCCATGATGACGCGCGGTC 900
242 alIeGluSerGlnCysTyrValIleAlaAlaGlnCysGlyArgHisH 259
      : :|||
901 TTATCGAAGACGGGCGCTGCTCCGCGCGCGGCAAGTGGCATC...G 947
259 IsGluThrArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 275
      :|||:| :|||
948 AGAGAACTGCATGCTGCTGGCCACTCTGCATGCTGGCGCGGCGG 997
276 ThrValValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAla 291
      :|||:| :|||
998 GAATGTGCTGCTCTCACTAGCAGCTGAGAGAGAGAGAGAGAGAGAG 1047
291 gIleAspLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe 307
      :|||:| :|||
1048 CGTCGATCTCGATCGCTCGCGGGAACACTGCGTGACACATCTTCACTTCA 1097
308 ..GlnHisArgArgProAspLeuTyrGly 316
1098 AGCAGCATGCTGACGCCCACTACTATGCT 1126

seq_name: /cgn2_6/ptodata/2/Ina/5A_COMB.seq:US-08-294-871A-5

seq_documentation_block:
: Sequence 5, Application US/08294871A
: Patent No. 5824522
: GENERAL INFORMATION:
: APPLICANT: Ikenaka, Yasuhiro
: APPLICANT: Nambu, Hirokazu
: APPLICANT: Takano, Masayuki
: APPLICANT: Yajima, Kazuyoshi
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
: APPLICANT: Yamada, Kazuhiko
: APPLICANT: Hiraiishi, Yoshiro
: TITLE OF INVENTION: Immobilized Enzyme Preparation and
: TITLE OF INVENTION: Process for Producing D-a-Amino Acid
: NUMBER OF SEQUENCES: 70
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/294,871A

```

```

: FILING DATE: 22-AUG-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,758
: FILING DATE: 12-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/917,111
: FILING DATE: 07-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/211,641
: FILING DATE: 11-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 400848/1990
: FILING DATE: 07-DEC-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP91/01696
: FILING DATE: 06-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 407922/1990
: FILING DATE: 27-DEC-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 078840/1991
: FILING DATE: 11-APR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 140051/1991
: FILING DATE: 12-JUN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP92/00739
: FILING DATE: 10-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 212692/1992
: FILING DATE: 10-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP93/01101
: FILING DATE: 05-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 340078/1992
: FILING DATE: 21-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Wegner, Harold C.
: REGISTRATION NUMBER: 25,258
: REFERENCE/DOCKET NUMBER: 74129/127/AOPA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 672-5300
: TELEFAX: (202) 672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1785 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: ORIGINAL SOURCE:
: STRAIN: JM109 PAD404 (FERM BP-3913)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 233..1144
: US-08-294-871A-5

alignment_scores:
: Quality: 208.00 Length: 343
: Ratio: 1.292 Gaps: 14
: Percent Similarity: 46.939 Percent Identity: 23.907

alignment_block:
US-09-357-675c-22 x US-08-294-871A-5 ..
Align seg 1/1 to: US-08-294-871A-5 from: 1 to: 1785

46 AlaValCysGlnValThr.....SerThrProAs 55
|||:||||| :|||
164 GCGATGTGTTTCACACGTTTCCCGCGCGCTGGCGGAGACATCACTGAG 213

```



```

55 nlysgIngluaSnPhelyThr..... 62
   |||:|||||:|||||
214 GAAGGAGCAGAGGTTTCATGACACGTCAGATGATTCTTCAGTGGGCAAC 263
63 .....Cys 63
264 AAGGTCCGATCCGCGCGGAGACACGCGAACAGTCCGTCTCTCT 313
64 ALaGluLeuValGInGluLaLaAatGLeuGlyAlaCysLeuAlaPheLe 80
   |||:|||||:|||||
314 CTCGACATGCTGACGAAGCCCGACCCGGGGCGCAATTCATTCTCT 363
80 uProGlu.....AlaPheAspPheLaLa 89
   |||:|||||
364 CCCCAGACTCCGCGCTTACGACCTTCTCCCGCGCTGCAATTCACCGAGC 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
   |||:|||||
414 AGCGCGAGCTGATGATCTTATGAGACCGAA...ATGCGCGCGCGG 460
106 LeuGlyGInTySerGInLeuAlaAatGlyCysGlyIleTrpLeuSerLe 122
   |||:|||||
461 GTCCGTCACCTCTTGAGAAAGCCGCGGAACTCGGATCGGCTTCAATCT 510
122 u.....GlyGlyPheHisGluAatGlyGln 131
   |||:|||||
511 GGGCTACGCTGAACTCGTCGAGAGCGCGCTCAAGCGTCG... 553
131 sPTRpGluGInAsnGInLysIleTyAsnCysHisValLeuAsnSer 147
   |||:|||||
554 .....TTCACACGTCGCAATTTGGTGCGATAG 580
148 LysGlySerValValAlaSerTyArgLysThrHisLeuCysAspValG 164
   |||:|||||
581 TCAGCGCAATGCTGCGCAAGATCGTAAGATCCATTTCG..... 619
164 uLepProGlyGln.....GlyProMetArg.....G 173
   |||:|||||
620 ....CCGGGTCAACAGAGTACGAGCCCTACCGCGCTTCACGATCTTC 665
173 LuserAsnTyThrLysProGlyGlyThrLeuGluProProValLysThr 189
   |||:|||||
666 AAAGCGTATTTGAGCGCGGATCTCGGCTCCCGGCTTACAGCTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyAspMetArgPheProG 206
   |||:|||||
716 GACGCGCGAATAATGGGATGTTTCATCTGCAACGATCGCGCTGGCTGA 765
206 uLeuSerLeuLysLeuAlaGluAlaGlyAlaGluIleLeuThrTyPro 223
   |||:|||||
766 ACCCTGGCGGGTATGGCTTCAGGGGCGCGAGATCCTGTC..... 808
223 eAlaPheGlySerValThrGlyProAlaHisTrpGluValLeu..... 237
   |||:|||||
809 .....GGCGGTACACACGCCGACCCACAATCCCTTGTTCGCCAG 850
238 .....LeuArgAlaArgAl 242
851 CACGACACCTGACGTCCTTCCACATCTCTATGATGACGAGCGCGGTC 900
242 AlLeuGluSerGInCysTyValIleAlaAlaAlaGInCysGlyArgHis 259
   |||:|||||
901 TTATTCAGAGCGGGCTGTCGCGCGCGGAGGTGGCATG...G 947
259 IsGluThrArgAlaSerTyGlyHisSerMetValValAspProTrpGly 275
   |||:|||||
948 AGGAGAACTGCATGCTGCTCGGCACTCCGTCATCTGGGCGCGAGCGG 997
276 ThrValValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAla 291
   |||:|||||
998 GAAATGTCGCTCTCACTACGACGCTGAAGAGAGAGGTGATCACCGCGG 1047

```

```

291 gIleAspLeuHisPheLeuGInImetArgGInHisLeuProValPhe. 307
   |||:|||||
1048 CGTGATCTGATCGCTGCGCGGAGACTGCTGAACACATCTTCACTTCA 1097
308 ..GInHisArgArgProAspLeuTyGly 316
   |||:|||||
1098 AGCAGCATGCTGACGCCGACACTATGCT 1126

seq_name: /cgn2_6/plodata/2/ina/6A_COMB.seq:US-08-876-398A-5
seq_documentation_block:
; Sequence 5, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKABA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25, 258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 PAD404 (FERM BP-3913)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-5

```

## alignment\_scores:

Quality: 208.00 Length: 343  
 Ratio: 1.292 Gaps: 14  
 Percent Similarity: 46.939 Percent Identity: 23.907

## alignment\_block:

US-09-357-675c-22 x US-08-876-398A-5 ..

Align seg 1/1 to: US-08-876-398A-5 from: 1 to: 1785

```

46 AlaValCysGlnValThr.....SerThrProAs 55
|||||:|||||:|||||:|||||:|||||:|||||:
144 GCGATGTGTTTCACACAGCTTTTCCCGCGCTGGCGCGACATCAGCTAG 213
|||||:|||||:|||||:|||||:|||||:|||||:
55 nLysGlnGluAsnPhelysThr..... 62
|||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTTCATGACACGTCATGATGACTTGACATGGGACAAAC 263
|||||:|||||:|||||:|||||:|||||:|||||:
63 .....Cys 63
264 AAGTCCGATCCGCGCGGAGACACGCGAAGAGTGTGCTGCTCTT 313
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80
|||||:|||||:|||||:|||||:|||||:|||||:
314 CTGCGACATGCTGACGAAAGCCGCGAGCCGCGCAATTTCATGCTCTT 363
|||||:|||||:|||||:|||||:|||||:|||||:
80 uProGlu.....AlaPheAspPheLeuAla 89
|||||:|||||:|||||:|||||:|||||:|||||:
364 CCCCAGAACCTCGGCTTACGACCTTCTCCCGCGCTGGCATTTCCACCGAGC 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
|||||:|||||:|||||:|||||:|||||:|||||:
414 AGCGGAGGCTGATAGCTTCTATGAGACCGAA...ATGCCGCGCGCGGTG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGlyCysGlyIleTyrLeuSerLe 122
|||||:|||||:|||||:|||||:|||||:|||||:
461 GTCCGCTCCACTCTTTGAGAGGCGCGGAACTCGGATCGGCTTCAATCT 510
122 u.....GlyIlePheHisGluArgGlyGln 131
|||||:|||||:|||||:|||||:|||||:|||||:
511 GGGCTACAGCTGAACTCGTCTGAAAGCGCGCGCTCAAGCTCTCC..... 553
131 sPTPrGluGlnAsnGlnIlyIleTyrAsnCysHisValLeuLeuAsnSer 147
|||||:|||||:|||||:|||||:|||||:|||||:
554 .....TTCACACAGTCATTTTGGTGATAG 580
148 LysGlySerValValAlaSerTyrArgIlyThrHisLeuCysAspValAl 164
|||||:|||||:|||||:|||||:|||||:|||||:
581 TCAGGCAAGATGCTGCGGCAAGTATGTAAGATCCATTTG..... 619
164 uLleProGlyGln.....GlyProMetArg.....G 173
|||||:|||||:|||||:|||||:|||||:|||||:
620 ...CCGGGTCAACAAAGAGTACGAGCGCTACCGCGCTTCCAGCATTTG 665
173 lUserAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
|||||:|||||:|||||:|||||:|||||:|||||:
666 AAAAGCGTATTTGAGCGCGGCGATCTCGGCTCCCGGTCAATAGCGTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
|||||:|||||:|||||:|||||:|||||:|||||:
716 GACGCGCGGAAATATGGGATGTTCTGCAACAGATCCGCCCTGGCGCTGA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyValGluLeuThrTyrProAs 223
|||||:|||||:|||||:|||||:|||||:|||||:
766 AGCCTGCGCGGATGAGGCGCTCAGGCGCGCGAGATCATCTCC..... 808
223 eTAlaPheGlySerValThrGlyProAlaHisThrGluValLeu..... 237
|||||:|||||:|||||:|||||:|||||:|||||:
809 .....GGCGGTACAAACAGCGCGACCCACATCCCTTGTCCCGAG 850
238 .....LeuArgAlaArgAl 242
|||||:|||||:|||||:|||||:|||||:|||||:

```

```

851 CACGACCACCTGACGTCCTTCCACCATCTCTATGATGACGCGCGGTC 900
242 atleGluSerGlnCysTyrValIleAlaAlaGlnCysGlyArgHis 259
|||||:|||||:|||||:|||||:|||||:|||||:
901 TTATCAGAAAGCGGCGCTGCTCCGCGCGCGCGCAAGTGGGATG...G 947
259 isGluThrArgAlaSerTyrGlyHisSerMetValValAspProTyrG 275
|||||:|||||:|||||:|||||:|||||:|||||:
948 AGGAGAACTGATGCTGCTGGCGCACTCTGCATCTGGCGCGCGAGCGG 997
276 ThrValValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAla 291
|||||:|||||:|||||:|||||:|||||:|||||:
998 GAATTCGTGCTCTCACTACAGCCTGGAAGACGAGTATCATCCGCGCG 1047
291 gLleAspLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe 307
|||||:|||||:|||||:|||||:|||||:|||||:
1048 CTGCGATCTCGATCGCTCGCGGGAAGCTGGCTGTAACATCTTCAACTTCA 1097
308 ..GlnHisArgArgProAspLeuTyrGly 316
|||||:|||||:|||||:|||||:|||||:|||||:
1098 AGCAGCATGCTCAGCCCGACAGCATATGATGT 1126

seq_name: /cgn2_6/ptodate/2/ina/5A_COMB.seq:US-08-294-871A-61
seq_documentation_block:
; Sequence 61, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanda, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraiishi, Yoshio
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990

```

```

; FILING DATE: 27-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/127/AOPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
;
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; ORIGINAL SOURCE:
; STRAIN: JM109 PAD468
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 233..1144
;
; US-08-294-871A-61

```

```

alignment_scores:
    Quality: 207.00      Length: 343
    Ratio: 1.286         Gaps: 14
    Percent Similarity: 46.939   Percent Identity: 23.907

```

```

alignment_block:
US-09-357-675C-22 x US-08-294-871A-61 ..

```

```

Align seg 1/1 to: US-08-294-871A-61 from: 1 to: 1785

```

```

46 AlavaLcysGlnValThr.....SerThrProAs 55
|||||:|||||:|||||:|||||:|||||:|||||:
164 GCGATGCTGTTCAACACGTTTCCCGCGCTGGCGGACATCACCCTAG 213
55 nLysGlnGluAsnPhelThr..... 62
|||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTCTATGACGTCAGATGATCTTGAGTGGGACAC 263
63 .....Cys 63
264 AAGTCCGATCGCGCGCGGAGACAGCGAACAGTCTGCTCTT 313
64 AlagLcyluValGlnGluAlaIaIaArgLcyluAlaCysLeuAlaPhe 80
|||||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATCTGACGAAGCCGCGAGCGCGGATTCATTCTT 363
80 uProGlu.....AlaPheAspPheIleAla 89
|||||:|||||:|||||:|||||:|||||:|||||:
364 CCCGAACTGCGCTTACGACATCTTCCGCGCTGACATTTCACGAG 413
89 rGAsnProAlaGluThrLeuLeuSerGluProLeuAsnGlyAspLeu 105
|||||:|||||:|||||:|||||:|||||:|||||:

```

```

414 AGCGGAGCTCGATAGCTCTATGACCGAA...ATGCCGCGCGCGGTG 460
106 LeuGlyInTySerGlnLeuAlaArgLcyluGlyLeuThrLeuSerLe 122
|||||:|||||:|||||:|||||:|||||:|||||:
461 GTCGTCACCTCTTGAAGAGCGCGGAACCTGCGATCGGCTTCAATCT 510
122 u.....GlyGlyPheHisGluArgGlyGln 131
|||||:|||||:|||||:|||||:|||||:|||||:
511 GGGGTACGCTGACCTGCTCGTCAAGCGCGGTACACGCTCC..... 553
131 sPTPrGluGlnAsnGlnLysIleTyraAsnCysHisValLeuLeuAsnSer 147
|||||:|||||:|||||:|||||:|||||:|||||:
554 .....TTCAACAGCTCATTTGGTGGATAG 580
148 LysGlySerValAlaIaSerTyraGlyThrHisLeuCysAspValG 164
|||||:|||||:|||||:|||||:|||||:|||||:
581 TCAAGCAAGATCGTCGCGCAAGTATCGTAATCCATTG..... 619
164 uLleProGlyGln.....GlyProMetArg.....G 173
|||||:|||||:|||||:|||||:|||||:|||||:
620 ....CCGGGTCAACAGAGTACGAGGCTTACCGGCTTCCAGCATCTTG 665
173 LuserAsnTyrrThrLysProGlyGlyThrLeuGluProProValLysThr 189
|||||:|||||:|||||:|||||:|||||:|||||:
666 AAAGCGTTATTTGACGCGCGGCGATCTCGGCTTCCGGTCTATGAGCTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyraAspMetArgPheProG 206
|||||:|||||:|||||:|||||:|||||:|||||:
716 GACCCCGCAAAATGGGATGTCATCTGCACACATGCGCGCTGCTGCA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyAlaIleLeuThrTyrrPro 223
|||||:|||||:|||||:|||||:|||||:|||||:
766 AGCTGCGCGGTGATGGCTCAGGCGCGCGGAGATATCTGC..... 808
223 eAlaPheGlySerValThrGlyProAlaHisTrpGluValLeu..... 237
|||||:|||||:|||||:|||||:|||||:|||||:
809 .....GGCGCTACACACGCGGACCCACATCCCGGTGTCGCCAG 850
238 .....LeuArgAlaArgAl 242
|||||:|||||:|||||:|||||:|||||:|||||:
851 CACGACCACTGACGCTTCCACACATCTCTATGATGACGCGCGGCTC 900
242 aLleGlySerGlnCysTyrrValIleAlaIaIaGlnCysGlnArgHis 259
|||||:|||||:|||||:|||||:|||||:|||||:
901 TTATCAGACAGGCGGCTGTCGCGCGCGCGGCGGAGGTGGCATG...G 947
259 IsGluThrArgAlaSerTyrrGlyHisSerMetValAlaAspProTrpGly 275
|||||:|||||:|||||:|||||:|||||:|||||:
948 AGGAGAACTGATCTGCTGCGCCACATCTCTGATCGTGGCGCGACCGGG 997
276 ThrValValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaAr 291
|||||:|||||:|||||:|||||:|||||:|||||:
998 GAATCTGCTGCTCTACTAGCAGCTGGAAGACGAGGTGATACACCGCGC 1047
291 gLleAspLeuHisPheLeuGlnIleMetArgGlnHisLeuProValPhe 307
|||||:|||||:|||||:|||||:|||||:|||||:
1048 CTTGATCTGATCTGCTGCGGGAAGTGGGAACATCTTCAACTTCA 1097
308 ..GlnHisArgArgProAspLeuTyrrGly 316
|||||:|||||:|||||:|||||:|||||:|||||:
1098 AGCAGCATCTGACCGCCAGCATATGTT 1126
seq_name: /cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-61
seq_documentation_block:
; Sequence 61, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAMBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio

```

```

1  APPLICANT: TAKAHASHI, SATOMI
2  TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
3  TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
4  NUMBER OF SEQUENCES: 70
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: FOLEY & LARDNER
7  STREET: 3000 K Street, N.W.
8  CITY: Washington
9  STATE: D.C.
10 COUNTRY: U.S.A.
11 ZIP: 20007-5109
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.30
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/876,398A
19 FILING DATE: 16-JUN-1997
20 CLASSIFICATION: 435
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 08/211,641
23 FILING DATE: 11-APR-1994
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: WO PCT/JP93/01101
26 FILING DATE: 05-AUG-1993
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: JP 340078/1992
29 FILING DATE: 21-DEC-1992
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: JP 212692/1992
32 FILING DATE: 10-AUG-1992
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Wegner, Harold C.
35 REGISTRATION NUMBER: 25, 258
36 REFERENCE/DOCKET NUMBER: 74129/130
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: (202) 672-5300
39 TELEFAX: (202) 672-5399
40 INFORMATION FOR SEQ ID NO: 61:
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 1785 base pairs
43 TYPE: nucleic acid
44 STRANDEDNESS: double
45 TOPOLOGY: linear
46 MOLECULE TYPE: DNA (genomic)
47 ORIGINAL SOURCE:
48 ORGANISM: Escherichia coli
49 STRAIN: JMI09 pad468
50 FEATURE:
51 NAME/KEY: CDS
52 LOCATION: join(233..1141)
53 US-08-876-398A-61

```

```

alignment_scores:
    Quality: 207.00      Length: 343
    Ratio: 1.266         Gaps: 14
Percent Similarity: 46.939      Percent Identity: 23.907

alignment_block:
US-09-357-675C-22 x US-08-876-398A-61 ..

Align seg 1/1 to: US-08-876-398A-61 from: 1 to: 1785

46 AlaValCysGlnValThr.....SerThrProAs 55
||||:|||| :|||:||||:||||:||||:
164 GCGATGCTGTTCACACAGCTTTCCCGCGCGCTGGCCGGACATCAGCTG 213
||||:||||:||||:||||:||||:||||:
55 nlysgIngluAsnphelysThr..... 62
||||:||||:||||:||||:||||:||||:
214 GAAGGAGCAAGGTCATGTGACACGTCAGATGATCTTGAGTGAGCAAC 263

```

[illegible]

|||||  
1098 AGCAGCATCTCAGCCAGCAGCATATGTT 1126

seq\_name: /g9n2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-294-871A-7

seq\_documentation\_block:

; Sequence 7, Application US/08294871A  
; Patent No. 5824522

; GENERAL INFORMATION:

; APPLICANT: Ikenaka, Yasuhiro

; APPLICANT: Namba, Hirokazu

; APPLICANT: Takano, Masayuki

; APPLICANT: Yajima, Kazuyoshi

; APPLICANT: Yamada, Yukio

; APPLICANT: Takahashi, Satomi

; APPLICANT: Okubo, Kazuma

; APPLICANT: Yamada, Kazuhiko

; APPLICANT: Hiraiishi, Yoshiko

; TITLE OF INVENTION: Immobilized Enzyme Preparation and

; TITLE OF INVENTION: Process for Producing D-a-Amino Acid

; NUMBER OF SEQUENCES: 70

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/294,871A

; FILING DATE: 22-APR-1994

; CLASSIFICATION: 435

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/971,758

; FILING DATE: 12-APR-1993

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/917,111

; FILING DATE: 07-AUG-1992

; APPLICATION DATA:

; APPLICATION NUMBER: US 08/211,641

; FILING DATE: 11-APR-1994

; APPLICATION DATA:

; APPLICATION NUMBER: JP 400848/1990

; FILING DATE: 07-DEC-1990

; APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP91/01696

; FILING DATE: 06-DEC-1991

; APPLICATION DATA:

; APPLICATION NUMBER: JP 407922/1990

; FILING DATE: 27-DEC-1990

; APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP92/00739

; FILING DATE: 10-JUN-1992

; APPLICATION DATA:

; APPLICATION NUMBER: JP 212692/1992

; FILING DATE: 10-AUG-1992

; APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP93/01101

; FILING DATE: 05-AUG-1993

; APPLICATION DATA:

; APPLICATION NUMBER: JP 340078/1992

; FILING DATE: 21-DEC-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Wegner, Harold C.

; REGISTRATION NUMBER: 25,258

; REFERENCE/DOCKET NUMBER: 74129/127/AOPA

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 672-5300

; TELEFAX: (202) 672-5399

; TELEX: 904136

; INFORMATION FOR SEQ. ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1785 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; ORIGINAL SOURCE:

; STRAIN: JM109 PAD406 (FERM BP-3914)

; NAME/KEY: CDS

; LOCATION: 233..1144

; US-08-294-871A-7

alignment\_scores:

Quality: 206.00 Length: 338

Ratio: 1.280 Gaps: 14

Percent Similarity: 47.633 Percent Identity: 24.260

alignment\_block:

US-09-357-675C-22 x US-08-294-871A-7 ..

Align seg 1/1 to: US-08-294-871A-7 from: 1 to: 1785

46 AAlaValCysGlnValThr.....SerThrProAs 55  
164 GCGATGTGTTTCACACGTTTTCGCCGCCGCTGGCGCGACATCATTAG 213  
55 nlysglnGlnuasnphelythr..... 62  
214 GAAGAGCAGAGGTTTCATGACACGTCAGATGATCTTCATGAGGACAC 263  
63 .....Cys 63  
264 AAGTCGATCGCGCGCGGAGACGCGAAGAGGTCGTCGTCGTT 313  
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80  
314 CTCGACATGCTGACGAAAGCCGCGAGCCGCGGCGGAAATTCATTGCTT 363  
80 uProGlu.....AlaPheAspPheIleAla 89  
364 CCGCGAAGCTCGGCTTACGACCTTCCCGCGCGCTGCGATTTCACGACG 413  
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105  
414 AGCGGAGCTCGATAGCTTATGAGACGAA..ATGCCGCGCGCGGTG 460  
106 LeuGlyGlnTyrSerGlnLeuAlaArgGluCysGlyIleTrrPleuSerLe 122  
461 GTCCGTCACCTTTTGAGAAAGCGCGGAACTCGGAGTTCGCTTCACT 510  
122 u.....GlyGlyPheHisGluArgGlyGln 131  
511 GCGCTACGCTGACTCGTCGCGAAGCGGCGTCAAGCGTGC..... 553  
131 sPrrpGluGlnAsnGlnLysIleTyrAsnGlyHisValLeuLeuAsnSer 147  
554 .....TTCAACAGCGTCATTTGGTGGAATAG 580  
148 LysGlySerValAlaLaserTyrArgLysThrHisLeuGlyAspValGln 164  
581 TCAGGCAAGATGCTGCGCAAGTATCGTAAGATCATTTG..... 619  
164 uLleProGlyGln.....GlyProMetArg.....G 173

```

620 .....|
173 luserstnrythlysproglygylthleuglupropovalylsthr 169
666 AAAAGCGTATTTCGAGCCGGCGATCTCGGCTCCGCTATGAGCTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProgl 206
716 GAGCGCGCGAAATGGGGATGTTCACTGCAACGATCGCGCGCTGA 765
206 uleuSerLeuLysLeuAlaGlnAlaGlyAlaIleLeuThr.....T 221
766 AGCGTCGGCGGTGATGGCCTCAGGGCGCGGAGATCATCTCGCGGCT 815
221 yProSerAlaPheGlySerValThrglyProAlaHstirpGluValLeu 237
816 ACAACAGCGCGACCCACAAATCCCTGTCTCCGACGACGACACTGACG 865
238 .....LeuArgAlaArgAlaIleGluSerGlnCy 247
866 TCCTTCACCATCTCTATGATGCAAGCGGGCTTATCAGAACGGGCG 915
247 stYrValIleAlaAlaAlaGlnCysGlyArgHstHsGluThrArgAla 264
916 CTGGTCGGCGCGCGGCAAGGTGGCATG...GAGAGAACTGCATGC 962
264 ertYrGlyHstSerMetValValAspProTrpGlyThrValAlaIla 280
963 TGCTCGCGCACTCTCGATCTGCGCGCGGCAAAATCGTCGCTC 1012
281 CysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLeuHst 296
1013 ACACAGAGCTGAGAACGAGGTGATCAGCGCGCGCTCATCTCGATCG 1062
296 eleuGlnGlnMetArgGlnHstLeuProValPhe...GlnHstArgArg 312
1063 CTGCCGGGAAGCTCGTGACACATCTTCACTTCAAGCAGCATCGTCAG 1112
312 roAspLeuArgly 316
1113 CCCAGCATATGAT 1126

seq_name: /cgn2_6/plodata/2/lna/5A_COMB.seq:US-08-294-871A-9

seq documentation block:
: Sequence 9, Application US/08294871A
: Patent No. 5824522
: GENERAL INFORMATION:
: APPLICANT: Ikenaka, Yasuhiro
: APPLICANT: Nambu, Hirokazu
: APPLICANT: Takano, Masayuki
: APPLICANT: Yajima, Kazuyoshi
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
: APPLICANT: Yamada, Kazuhiko
: APPLICANT: Hiraiishi, Yoshiro
: TITLE OF INVENTION: Immobilized Enzyme Preparation and
: NUMBER OF SEQUENCES: 70
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/294,871A
: FILING DATE: 22-AUG-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,758
: FILING DATE: 12-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/917,111
: FILING DATE: 07-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/211,641
: FILING DATE: 11-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 400848/1990
: FILING DATE: 07-DEC-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP91/01696
: FILING DATE: 06-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 407922/1990
: FILING DATE: 27-DEC-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 078840/1991
: FILING DATE: 11-APR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 140051/1991
: FILING DATE: 12-JUN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP92/00739
: FILING DATE: 10-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 212692/1992
: FILING DATE: 10-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP93/01101
: FILING DATE: 05-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 340078/1992
: FILING DATE: 21-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Wegner, Harold C.
: REGISTRATION NUMBER: 25,258
: REFERENCE/DOCKET NUMBER: 74129/127/AOPA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 672-5300
: TELEFAX: (202) 672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1785 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: ORIGINAL SOURCE:
: STRAIN: JM109 PAD416 (FERM BP-3915)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 233..1144
: US-08-294-871A-9

alignment_scores:
: Quality: 206.00 Length: 343
: Ratio: 1.280 Gaps: 14
: Percent Similarity: 46.939 Percent Identity: 23.907

alignment block:
US-09-357-675c-22 x US-08-294-871A-9 ..
Align seg 1/1 to: US-08-294-871A-9 from: 1 to: 1785
46 AlaValCysGlnValThr.....SerThrProAs 55

```







```

? COUNTRY: U.S.
? ZIP: 20007-5109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/294,871A
? FILING DATE: 22-AUG-1994
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/971,758
? FILING DATE: 12-APR-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/917,111
? FILING DATE: 07-AUG-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/211,641
? FILING DATE: 11-APR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 400848/1990
? FILING DATE: 07-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP91/01696
? FILING DATE: 06-DEC-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 407922/1990
? FILING DATE: 27-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 078840/1991
? FILING DATE: 11-APR-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 140051/1991
? FILING DATE: 12-JUN-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP92/00739
? FILING DATE: 10-JUN-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 212692/1992
? FILING DATE: 10-AUG-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP93/01101
? FILING DATE: 05-AUG-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 340078/1992
? FILING DATE: 21-DEC-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Wegner, Harold C.
? REGISTRATION NUMBER: 25,258
? REFERENCE/DOCKET NUMBER: 74129/127/AOPA
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 672-5300
? TELEFAX: (202) 672-5399
? TELEEX: 904136
? INFORMATION FOR SEQ ID NO: 21:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1785 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? ORIGINAL SOURCE:
? STRAIN: JM109 PAD439
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 233..1144
? US-08-294-871A-21

```

```

alignment_scores:
  Quality: 206.00      Length: 343
  Ratio: 1.280        Gaps: 14
  Percent Similarity: 46.939  Percent Identity: 23.907

```

```

alignment_block:
US-09-357-675c-22 x US-08-294-871A-21 ..
Align seg 1/1 to: US-08-294-871A-21 from: 1 to: 1785

46 AlAValcysGlnValThr.....SerThrProAs 55
164 GCGATGTGTTTCAACACGTTTTCGGCGCGCTGGCGCGACATCAGCTAG 213
55 nlySGInGluasnphelyThr..... 62
214 GAAAGACAGAGGTTTCATGACAGTCAGATGATCTTGCAGTGGGACAAAC 263
63 .....Cys 63
264 AAGTCCGATCGCGCGCGAGACAGCGCAAGCTGCTGCTGCTT 313
64 AlaGluLeuValGInGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
314 CTCGACATGCTGACGAAAGCCGCGAGCGCGCGCATTTTCATTGCTT 363
80 uProGlu.....AlaPheAspPheIleAla 89
364 CCCGAACTCGCGCTTACGACCTTCTCCGCGCTGCGATTTCACGACG 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
414 AGCGCGAGCTCGATAGCTTCTATGACCGAA...ATGCCGCGCGGTG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGluCysGlyIleTyrLeuSer 122
461 GTCCGTCACACTTTGAGAAAGCCGCGGAACTCGGATCGGCTTCATCT 510
122 u.....GlyGlyPheHisGluArgGlyGln 131
511 GGGCTACGCTGAACTCGTCTCGAAGCGCGGTCGAAGGTCG..... 553
131 sPTpGluGlnAsnGlnLysIleTyrAsnGlyHisValLeuLeuAsnSer 147
554 .....TTCANACGCTCATTTTGTGATGAG 580
148 LysGlySerValAlaAlaSerTyrArgLysThrHisLeuGlyAspValG 164
581 TCAGGCAAGATCGTCGCAATGATCGTAAGATTCATTG..... 619
164 uIleProGlyGln.....GlyProMetArg.....G 173
620 ....CCGGTCCACAAGAGTACGAGGCGCTACCGCGCTTCCAGATCTTG 665
173 LuserAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
666 AAAAGCGTTATTTCGAGCCGGCGCATCTCGGCTTCCCGGTATGACGTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
716 GAGCCCGGAAATGGGATGTCATGCAACGATCCGCGCGCTGCA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 223
766 AGCCTGGCGGGTATGGGCTCAGGGCGCGCATCATCTGCG..... 808
223 eAlaPheGlySerValThrGlyProAla..... 232
809 .....GGCGGCTACACAGCGCGACCCACATCCCGCTTCCCGAG 850
233 .....HisTyrGluValLeuLeuArgAlaArgAl 242
851 CAGCACACCTGACGTCCTTCCACCATCTCTATGATGACGCGGGGTC 900
242 aIleGluSerGlnCysTyrValIleAlaAlaGlnCysGlyArgHisH 259
901 TTATCAGAAAGCGGCTGCTGCGCGCGCGGCAAGGCTGCGATG...G 947

```

```

259 1sgluThrArgAlaSerTyrGlyHisSerMetValAlaSprTrpGly 275
    ::::: ||||| ::::: |||
948 AGAGAAACGATCGTCGCGGCGACTCCGATCGTGGCGGCGACGG 997
    ::::: ||||| ::::: |||
276 ThrValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAla 291
    ::::: ||||| ::::: |||
998 GAATGTGCTCTCCTCAGTACGAGCTGTAAGACGAGTATCACCCCG 1047
    ::::: ||||| ::::: |||
291 gltAspLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe 307
    ::::: ||||| ::::: |||
1048 CGTCGATCTCGATCGTCCGCGGAGTGGTGAACACATCTTCACTCA 1097
    ::::: ||||| ::::: |||
308 ..GlnHisArgArgProAspLeuTyrGly 316
    ||||| ::::: |||||
1098 AGCAGCATGTCGACGCCCAAGCATATG 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-59

seq_documentation_block:
: Sequence 59, Application US/08294871A
: Patent No. 5824522
:
: GENERAL INFORMATION:
: APPLICANT: Ikemaka, Yasuhiro
: APPLICANT: Namba, Hirokazu
: APPLICANT: Takano, Masayuki
: APPLICANT: Yajima, Kazuyoshi
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
: APPLICANT: Yamada, Kazuhiko
: APPLICANT: Hiratschi, Yoshiro
: TITLE OF INVENTION: Immobilized Enzyme Preparation and
: TITLE OF INVENTION: Process for Producing D-a-Amino Acid
: NUMBER OF SEQUENCES: 70
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.
: ZIP: 20007-5109
:
: COMPUTER-READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/294,871A
: FILING DATE: 22-AUG-1994
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,758
: FILING DATE: 12-APR-1993
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/917,111
: FILING DATE: 07-AUG-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/211,641
: FILING DATE: 11-APR-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 400848/1990
: FILING DATE: 07-DEC-1990
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP91/01696
: FILING DATE: 06-DEC-1991
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 407922/1990
: FILING DATE: 27-DEC-1990
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 078840/1991
: FILING DATE: 11-APR-1991
:
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: JP 140051/1991
: FILING DATE: 12-JUN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP92/00739
: FILING DATE: 10-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 212692/1992
: FILING DATE: 10-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP93/01101
: FILING DATE: 05-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 340078/1992
: FILING DATE: 21-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Wegner, Harold C.
: REGISTRATION NUMBER: 25,258
: REFERENCE/DOCKET NUMBER: 74129/127/AOPA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 672-5300
: TELEFAX: (202) 672-5399
: TELEX: 904136
:
: INFORMATION FOR SEQ ID NO: 59:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1785 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: ORIGINAL SOURCE:
: STRAIN: JM109 PAD456
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 233..1144
: US-08-294-871A-59

```

```

alignment_scores:
    Quality: 206.00      Length: 342
    Ratio: 1.280         Gaps: 15
    Percent Similarity: 47.076      Percent Identity: 24.854

```

alignment\_block:

US-09-357-675c-22 x US-08-294-871A-59 ..

Align seg 1/1 to: US-08-294-871A-59 from: 1 to: 1785

```

46 AlaValCysGlnValThr.....SerThrProAs 55
    ||::: ||| ::::: |||
164 GCGATGTGTTCCACAACGTTTCCCGCGCGTGGCGGAGACATCACCTAG 213
    ::::: ||||| ::::: |||
55 nlYsgIngluAsnPhelysThr..... 62
    ::::: ||||| ::::: |||
214 GAAGGACGAGAGGTTTCATGACACGTCAGATGATCTTCAGTGGGACAC 263
    ::::: ||||| ::::: |||
63 .....Cys 63
264 AAGGTCGATCGCGCGCGGAGACACGCGAAGAGTGTGTCGTCCTT 313
    ::::: ||||| ::::: |||
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
    ::::: ||||| ::::: |||
314 CTCGACATCTGTCAGCAAGCCCGGAGCCGGGCGGAATTGATTTGTCCTT 363
    ::::: ||||| ::::: |||
80 uproGlu.....AlaPheAspPheIleAlaArg.....Asp 91
    ::::: ||||| ::::: |||
364 CCGCGAATCGCGCTTACGACCTTCTTCCCGCGCTGTGATTTTCACCGAG 413
    ::::: ||||| ::::: |||
91 roAlaGluThrLeuLeuLeuSerGlu...ProLeuAsnGlyAspLeuLeu 106
    ||||| ::::: ||| ::::: |||
414 AGGCGAGCTCGATAGCTTCTATAGAGACGGAATGCGCGCGGTGTC 463
    ::::: ||||| ::::: |||
107 GlyGluTyrSerGlnLeuAlaArgGluCysGlyIleTrpLeuSerLeu.. 122
    ::::: ||| ::::: |||
464 CTCACATCTTTGAGAAAGCCCGGAACATCTGGGATCGCTTCATCTGGG 513

```

```

123 ..... GlyGlyPheHisGluArgGlyGlnAspT 132
514 CTACGCTGAACTCGTGCTCGAGAGCGGCTCAAGCGCTGC..... 553
132 rpcluglnaenglulyslletryasnucyshsvalleuleuasnserlys 148
554 ..... TTCACACGTCATTTTGGTGATTAAGTCA 583
149 GlySerValAlaIasertyrarglysthisleucysaspyalglu1 165
584 GGCAGAGTCGTGCGCAAGTATCGTAAGATCCATTG..... 619
165 eproglygln..... GlyPrometArg..... Glus 174
620 .CGGGTCACAAAGAGTACGAGCGCTACCGCGCTCCACACTTGTAA 668
174 etasntrythrlystproglyglythrleuglupropovalysthrpro 190
669 AGCGTATTTCGAGCGCGGCGATCTCGGCTCCGGCTATGACGTGAC 718
191 AlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProGlu 207
719 GCCGCGAAATGGGATGTTCATCTGCAACGATCGCGCTGCCCTGAAGC 768
207 usertleuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProsera 224
769 CTGCGCGGTCATGGGCTCAGAGCGCGCGAGATCATCTGC..... 808
224 laPheGlySerValThrGlyProAlaHisTrpGluValLeu..... 237
809 ..... GCGGCTACAAACACGCCGCCACCAATCCCTGTGTTCCGCCACAC 853
238 ..... LeuArgAlaArgAla1 243
854 GACCACCTGACGTCCTTCACACATCTCTATGATGACGCGCGGCTCTA 903
243 egluserGlnCysetyrValIleAlaAlaIaGlnCysGlyArgHisHisG 260
904 TCAGAAAGGGGCTGTGCTCGCGCGCGGCAAGTCAGGCAATG...GAGG 950
260 luthrArgAlaIasertyrGlyHisSerMetValValAspProTrpGlyThr 276
951 AGAAGTCGATGCTGCTGCGGCACTCCCTGCAATCGTGGCGCGGCGGAA 1000
277 ValValAlaIaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArg1 292
1001 ATCGTCGCTCTCACTAGACGCTGGAAGACGAGGTGATCACGCGCGCT 1050
292 eaSPLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe...G 308
1051 CGATCTCGATCGCTGCGGGAAGTGTGTAACACATCTTCAACTTCAAGC 1100
308 lnhIstArgArgProAspLeuTyrGly 316
1101 AGCATGCTCAGCCCGACACTATGCT 1126

seq_name: /cgn2.6/ptdata/2/lna/6A.COMB.seq:US-08-876-398A-7
seq_documentation_block:
; Sequence 7, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAMBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Megner, Harold C.
; REGISTRATION NUMBER: 25, 258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 PAD406 (FERM BP-3914)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-7

alignment_scores:
Quality: 206.00 Length: 338
Ratio: 1.280 Gaps: 14
Percent Similarity: 47.633 Percent Identity: 24.260

alignment_block:
US-09-357-675C-22 x US-08-876-398A-7 ..

Align seg 1/1 to: US-08-876-398A-7 from: 1 to: 1785

46 AlaValCysGlnValThr.....SerThrProAs 55
||||:||||:||||:||||:||||:||||:||||:
164 GCGATGTGTTTCACACACGTTTTCGCGCGCTGGCGGAGCATATCCTAG 213
55 nlysglnglnuasnpheLysThr..... 62
||||:||||:||||:||||:||||:||||:
214 GAAGAGACAGAGTTTCATGACAGCTCAGATGATCTTGAGTGGAGACAAC 263
63 .....Cys 63
264 AAGTCCGATCGCGCGGAGAGACAGCGAAGTGTGCTGCTCTT 313
64 AlaGluLeuValGlnGluAlaIaArgLeuGlyAlaCysLeuAlaPheLe 80

```

314 CTCGACATCTGACGAAACCGGACCGGCGCGAATTTCAATTCCT 363  
80 UPProGlu.....AlpheasppheleAla 89  
364 CCCCCAATCGCGCTTACGACCTTCTCCCGCGCTGGCATTTACCGAG 413  
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105  
414 AGCCGAGCTCGATAGCTTCTATGAGACCGAA...ATGCCGCGCCGGTG 460  
106 LeuGlyGlnTyrSerGlnLeuAlaGlyGlyIleTyrLeuSer 122  
461 GTCCTGCACCTTTTGAGAGCGCGGGAACGCGGATCGCTCAATCT 510  
122 U.....GlyGlyPheHisGluArgGlyGln 131  
511 GGGCTACGCTGAACCTGCTCGACAGCGCGCTCAAGCTCGC..... 553  
131 SPTrpGluInAsnGlnTyrAsnCysHisValLeuLeuAsnSer 147  
554 .....TTCACACGTCATTTTGGTGATAG 580  
148 LysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspValG 164  
581 TCAGGCAAGATCGTCGCAAGTATCGTAGATCCATTTG..... 619  
164 uIlePrgLysGln.....GlyPrometArg.....G 173  
620 ...CGGGTTCACAGAGATACGAGCCCTACCGCGCTTCACGATCTTG 665  
173 LysAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189  
666 AAAAGCGTATTTCAGCGCGGCGATCTCGCGTCCCGGTCATAGACGTC 715  
190 ProAlaGlyValGlyLeuAlaIleCysTyrAspMetArgPheProG 206  
716 GACGCGCGCAAAATGGGATGTTCATCTGCACGATCGCGCGGCTGA 765  
206 uLeuSerLeuLysLeuAlaGlnAlaGluIleLeuThr.....T 221  
766 AGCCTGCGCGGATGATGGCTCAGGGGCGCCGAGATCATCTGCGCGCT 815  
221 TyrProSerAlaPheGlySerValThrGlyProAlaHisTrpGluVal 237  
816 ACAACACGCGGACCCACAAATCCCTGTCCCGACGACGACGACGAGC 865  
238 .....LeuArgAlaArgAlaIleGluSerGln 247  
866 TCCTTCACACATCTCTATGATGACGCGCGGCTTATCAGAAAGCGGC 915  
247 sTyValIleAlaAlaGlnCysGlyArgHisHisGluThrArgAla 264  
916 CTGGTCCGCGCGCGCGGCAAGTGGCGATG...GAGGAGAACTGCATGC 962  
264 exTyrGlyHisSerMetValValAspProTyrGlyThrValValAla 280  
963 TCGTCGCGCAATCTCGCATCGTCGCGCGGCAAGTTCGCTTC 1012  
281 CysSerGln...GlyProGlyLeuCysLeuAlaArgIleAspLeuHis 296  
1013 ACTACGACCGCTGAGAACGAGGATCACCGCGCGCTGATCTCGATCG 1062  
296 eleuGlnGlnMetArgGlnHisLeuProValPhe...GlnHisArgArg 312  
1063 CTGCGCGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATGTCAGC 1112  
312 roAspLeuTyrGly 316  
1113 CCCAGCATATGTT 1126

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-9

seq\_documentation\_block:  
; Sequence 9, Application US/08876398A  
; Patent No. 6083752  
; GENERAL INFORMATION:  
; APPLICANT: IKENAKA, Yasuhiro  
; APPLICANT: NAMBA, Hirokazu  
; APPLICANT: TAKANO, Masayuki  
; APPLICANT: YAJIMA, Kazuyoshi  
; APPLICANT: YAMADA, Yukio  
; APPLICANT: TAKAHASHI, Satomi  
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/876,398A  
; FILING DATE: 16-JUN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/211,641  
; FILING DATE: 11-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP93/01101  
; FILING DATE: 05-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 340078/1992  
; FILING DATE: 21-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 212692/1992  
; FILING DATE: 10-AUG-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wegner, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 74129/130  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1785 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: JM109 pad416 (FERM BP-3915)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(233..1141)  
US-08-876-398A-9

alignment\_scores:  
Quality: 206.00 Length: 343  
Ratio: 1.280 Gaps: 14  
Percent Similarity: 46.939 Percent Identity: 23.907

alignment\_block:

US-09-357-675c-22 x US-08-876-398A-9 ..

Align seg 1/1 to: US-08-876-398A-9 from: 1 to: 1785



---

OM of: us-09-357-675c-25 to: GenEmbl:\* out\_format : pfs  
Date: Apr 29, 2002 11:15 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=framer.p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09357675/runat\_29042002\_091757\_3179/app-query.fasta.1.1208  
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.500 -OGAPEXT=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum2  
-TRANS=human40.ccd1 -LIST=45 -DCCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US09357675.ecgn1.1.10065 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPRXY -WAIT -THREADS=1

## Search information block:

Query: US-09-357-675C-25  
Query length: 362

Database: GenEmbl:\*

Database sequences: 1797656

Search length: 1873333701

Search time (sec): 3374.980000

## score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pr:AF069987	+ 1836.00	2332.76	1.4e-121	1385	AF069987 Homo sapiens nitrilas
gb_hcg:AL354714-4	+ 1538.50	1915.16	2.6e-98	110000	Continuation (5 of 7) of AL3
gb_hcg:AL354714-5	+ 1538.50	1915.16	2.6e-98	110000	Continuation (6 of 7) of AL3
gb_hcg:AL590651	- 1538.50	1911.42	4.2e-98	167863	AL590651 Homo sapiens chrom
gb_pr:AL591806	+ 1538.50	1909.84	5.1e-98	200822	AL591806 Human DNA sequence
gb_pr:AF069984	+ 1525.50	1927.70	5.2e-99	4079	AF069984 Homo sapiens nitrilas
gb_ro:AF069988	+ 1496.00	1899.97	1.8e-97	1338	AF069988 Mus musculus nitrilas
gb_ro:BC021634	+ 1478.00	1876.86	3.5e-96	1365	BC021634 Mus musculus, nitrila
gb_hcg:AC084821	+ 1105.00	1357.80	2.9e-67	211771	AC084821 Mus musculus chrom
gb_ro:AF069985	+ 1103.00	1388.07	5.5e-69	4481	AF069985 Mus musculus nitrilas
gb_hcg:AC087229	+ 1083.00	1329.08	1.1e-65	213203	AC087229 Mus musculus chrom
gb_hcg:AC105589	+ 1064.00	1306.45	2.1e-64	178418	AC105589 Rattus norvegicus c
gb_ov:AF284575	+ 968.50	1228.88	4.4e-60	13184	AF284575 Xenopus laevis Nit pr
gb_hcg:AC017755	+ 715.50	865.19	7.9e-40	1215	AC017755 Drosophila melanog
gb_in:AC093121	- 715.50	863.04	1.0e-39	168250	AC093121 Drosophila melanog
gb_in:AC093501	- 715.50	862.77	1.1e-39	173494	AC093501 Drosophila melanog
gb_hcg:AC006169	+ 715.50	862.73	1.1e-39	174279	AC006169 Drosophila melanog
gb_in:AC003467	+ 715.50	861.95	1.2e-39	190351	AC003467 Drosophila melanog
gb_in:AF069989	+ 710.50	857.97	2.0e-39	298640	AF069989 Drosophila melanog
gb_in:AF069986	+ 702.00	888.24	4.1e-41	1385	AF069986 Caenorhabditis eleg
gb_pr:AF372804	+ 648.50	821.56	2.1e-37	1172	AF372804 Arabidopsis thaliana
gb_hcg:AC010013	- 632.50	771.43	1.3e-34	34013	AC010013 Drosophila melanog
gb_ba:AE000934	- 537.50	657.78	2.9e-28	15124	AE000934 Methanobacterium the
gb_pl:SCUJ126W	+ 519.00	655.64	3.7e-28	1307	249401 S.cerevisiae chromosome
gb_pl:AF284571	+ 519.00	653.90	4.6e-28	1591	AF284571 Saccharomyces cerevis
gb_ba:AE000461	- 502.00	615.61	6.3e-26	10466	AE000461 Pseudomonas aerugin
gb_pl:SPBC651	+ 502.00	607.79	1.7e-25	25350	AE005572 S.pombe chromosome 1
gb_pr:AF260334	- 493.00	624.72	2.0e-26	1019	AF260334 Homo sapiens CUX002 m
gb_pl:SCGCD14	- 493.00	615.82	6.1e-26	2792	254149 S.cerevisiae GC014 gene
gb_ba:SYCSLND	- 490.00	579.47	6.5e-24	110908	D64002 Synecocystis sp. PCG
gb_pr:BC020620	+ 487.50	618.19	4.5e-26	965	BC020620 Homo sapiens Nit prot
gb_pr:AF284574	+ 485.50	615.67	6.2e-26	965	AF284574 Homo sapiens Nit prot
gb_ba:AE005512	- 481.00	600.02	4.7e-25	3180	AF000552 Nostoc commune trna-tl
gb_ro:AF284573	- 481.00	607.34	1.8e-25	1282	AF284573 Mus musculus Nit prot
gb_ba:AB017194	+ 479.50	593.90	1.0e-24	4762	AB017194 Plectonema boryanum C
gb_ro:BC020153	+ 479.00	606.51	2.0e-25	1063	BC020153 Mus musculus, RIKEN C
gb_ba:AP003587	+ 477.50	553.56	1.8e-22	343550	AP003587 Nostoc sp. PCC 7120
gb_ba:AL646071	+ 471.50	550.56	2.6e-22	203050	AL646071 Ralstonia solanaceae
gb_sts:G72919	+ 471.00	598.33	5.8e-25	847	G72919 MARC 2849-2850:991933517

gb\_ba:SME591791 - 444.50 511.59 3.9e-20 340900 ! AL591791 Sinorhizobium me  
gb\_ba:AJ414158 + 440.50 509.78 4.9e-20 235050 ! AJ414158 Yersinia pestis  
gb\_ba:AE004129 + 436.50 531.83 2.9e-21 10881 ! AE004129 Vibrio cholerae c  
gb\_ba:AE005759 - 433.50 526.61 5.7e-21 12748 ! AE005759 Caulobacter cresc  
gb\_pl:SPAC26A3 + 430.50 513.10 3.2e-20 38149 ! Z69240 S.pombe chromosome

seq\_name: gb\_pr:AF069987

seq\_documentation\_block:

LOCUS AF069987 1385 bp mRNA linear PRI 23-JUL-1998

DEFINITION Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.

ACCESSION AF069987

VERSION AF069987.1 GI:3228665

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

17 uargMetValleuallaleeserCysArgThTySerLeuSerArga 34

|||||  
53 GCCAATGGTTTGGCTATATCTTCATGCTG..... 82  
34 rrpProArgluLeuArgPheleThrArgProProHisArghPheleuSerleu 50  
83 .....GGCTTCATCCACAGGCGCTCTCACAGATTCCTGTCCTT 121  
51 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67  
122 CTTGTCCTGAGACTCCGATACCTCAACTCTCAGTACTTGTCTCAGCC 171  
67 oArgProArghAlaMetAlaIleSerSerSerSerCysGlnLeuProleu 84  
172 CAGGCCAGAGCCATGCTATCTCTCTCTCTCTCGAACTGCCCTGG 221  
84 aAlaValCysGlnValThrSerThrProAspLysGlnIleAsnPhel 100  
222 TGGCTGTGTGCCAGTACATGCAGCCAGACAGACAGAACTTTAA 271  
101 ThrCysAlaGluLeuValArgGluAlaIleArgLeuGlyAlaCysleuAl 117  
272 ACATGTGCTGAGCTGCTCGAGAGCTGCCAGACTGGTCCCTGCTGGC 321  
117 aPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrL 134  
322 TTTCTGCTGAGGCAATTGACTTCAATGCAGCGACCTGCAGAGAGCG 371  
134 euHisLeuSerGluProLeuGlyGlyLysLeuLeuGluIuTyrrThrl 150  
372 TACACCTGTCTGAACACACCTGGGGAACTTTGGAAAGATACACCAG 421  
151 LeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluAr 167  
422 CTTGCCAGGAAATGTGACTCTGCTGCTGCTGGTGGTTCATGAGCG 471  
167 gGlyIleAspTrpGluGlnThrGlnLysIleTyrrAsnGlyHisValLeuL 184  
472 TGGCCAGAGCTGGAGACAGACTGAAATCTAACAATTGTCAGCTCTCG 521  
184 euAsnSerLysGlyAlaValAlaIleThrTyrrArgLysThrHisLeu 200  
522 TGAACAGCAAGGGGAGTAGTGGCCACTTACAGGAAGACACACTGTGT 571  
201 AspValGluIleProGlyGlnGlyPrometCysGluSerAsnSerThr 217  
572 GACGTAGAGATTCCAGGCGAGGGCTATGTGAAAGCAACTCTACAT 621  
217 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 224  
622 GCTTGGGCCAGCTTGAGCTCACCTGTCAACACACAGCAGGCAAGATT 671  
234 lLeuAlaValCysTyrrAspMetArgPheProGluLeuSerLeuAlaLeu 250  
672 GTCTACTCTGTGCTATGACATGCGGTTCCTCAACTCTCTCGGCAAT 721  
251 AlAGlnAlaGlyAlaGluIleLeuThrTyrrProSerAlaPheGlySer 267  
722 GCTCAAGCTGGAGCAGAGATTAATCTTCCCTTTCAGCTTTTGGATCC 771  
267 eThrglyProAlaHisTrpGluValLeuLeuArgAlaIleGluT 284  
772 TACAGGCCAGGCCACTGGGAGGTGTGCTGGGGCCCGTGTATGAA 821  
284 hGlnCysTyrrValAlaAlaAlaGlnCysGlyTyrrGlnHisGluLys 300  
822 CCCAGGTCTATGATGAGCAGCAGCAGCACTGTGAGCCACCACTGGAAG 871  
301 ArgAlaSerTyrrGlyHisSerMetValAlaAspProTrpGlyThrVal 317  
872 ACAGCAAGTTATGGCCACAGCATGTGGTGAACCCCTGGGGAACAGTGG 921  
317 lAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 334  
|||||

922 GCGCCGCTGCTGAGGGGCCAGGCGCTTGCCTTGCCCGAATAGACCTCA 971  
334 sntYrLeuArghGlnLeuArghIleLeuProValPheGlnHisArgArg 350  
972 ACTATCTGCACAGTGTGGCCGACACCTGCTGTGTCCAGCACCGCAG 1021  
351 ProAspLeuTyrrGlyAsnLeuGlyHisProLeuSer 362  
1022 CCTGACCTCTATGCAATCTGGGTCAACCACTGTCT 1057  
seq\_name: gb\_htg:AL354714\_4  
seq\_documentation\_block:  
WPCOMMENT  
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714  
Fragment Name Begin End  
AL354714\_0 1 110000  
AL354714\_1 100001 210000  
AL354714\_2 200001 310000  
AL354714\_3 300001 410000  
AL354714\_4 400001 510000  
AL354714\_5 500001 610000  
AL354714\_6 600001 699820  
Continuation (5 of 7) of AL354714 from base 400001 (AL354714 Homo sapiens chromosome  
alignment\_scores:  
Quality: 1538.50 Length: 888  
Ratio: 4.262 Gaps: 9  
Percent Similarity: 40.653 Percent Identity: 40.315  
alignment\_block:  
US-09-357-675C-25 x AL354714\_4 ..  
Align seg 1/1 to: AL354714\_4 from: 1 to: 110000  
1 ProLeuAlaAlaAla\*\*LeuAlaProAspArgProProAspArgThrIle 17  
|||||  
103174 CCACGCTGCGCGCGCTTGGCTCCAGACCGCCCTCCGATCGGACCT 103223  
17 uArgMetValLeuAlaIleSerSerCysArghTrpTyrrSerLeuSerArg 34  
|||||  
103224 GCGAATGTTTGGCTATATCTTATGATGAGACTACTCCATATCCGCT 103273  
34 rrpPro..... 35  
103274 GCGCGGCGTGAATCCACCTCGCGTCTTTAACTGTGTAACAGAGATG 103323  
35 ..... 35  
103324 TGCCTCTGGAGAGCGCGGAGGAGGAGGCACTGAGAGCGGCGCGG 103373  
35 ..... 35  
103374 AGGCTGAGAGCGCGGCGGCTTGGGCTTGGGTTCTTTGCCCTTGC 103423  
35 ..... 35  
103424 CCACAGGAGGCGGTGGAGACGAGAGAGGAGGTAATTTCCTCGCAGA 103473  
35 ..... 35  
103474 TTCTGTGAAGGGAATATGCTTTCAGTCAAGTAAGCTGCGCAAGTGC 103523  
35 ..... 35  
103524 ACAGTCAAGAGAGAGTCTTGGAAACCAAGAGATGTTCCCGGAGATGA 103573  
35 ..... 35  
103574 CTTTGTGACTGCGGAAGCTTTGTCAAGAGAAAGAGCTTCAAGTTAATG 103623  
35 ..... 35





276 ..... 276  
105474 TAGTCACATGGGTAGATTGCTCTGTAATGTCCCTCACCTGCTACCTCC 105523  
276 ..... 276  
105524 CACTATTGCTACATGTACTTAAGTGAACACACATCTCATGCCAGGTGT 105573  
277 .. LeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAla 292  
105574 TGTGTGGGGGCGGCTGCTATCGAAACCAAGTGTATGTAGTGGACGACCA 105623  
293 GlnCysGlyArgHisHisGluValArgAlaSerTyrGlyHisSerMetVal 309  
105624 CAGTGTGACGCGCACCATGAGAGAGCAAGTTATGGCCACAGCATGGT 105673  
309 ValAspProTyrPglYthrValValAlaArgCysSerGluGlyProGlyL 326  
105674 GGTAAGCCCTGGGGAAAGTGTGGCCGCTCTGAGGGGCGCAGGCC 105723  
326 eucCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 342  
105724 TCTGCTTGCCTGCGAATGACCTCACTATCTGCGACAGTTGCCCGACAC 105773  
343 LeuProValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGlyHis 359  
105774 CTGGCTGTCTTCAGACCGCAGCGCCTGACCTCTATGGCAATCTGGGTCA 105823  
359 sProLeuSer 362  
105824 CCCACTGTCT 105833  
seq\_name: gb\_htg:AL354714\_5

seq\_documentation\_block:  
WPCOMMENT

Sequence split into 7 fragments LOCUS AL354714 Accession AL354714  
Fragment Name Begin End  
AL354714\_1 1 110000  
AL354714\_2 110001 210000  
AL354714\_3 210001 310000  
AL354714\_4 310001 410000  
AL354714\_5 410001 510000  
AL354714\_6 510001 610000  
AL354714\_7 610001 699820  
Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome 1 c

alignment\_scores:  
Quality: 1538.50 Length: 888  
Ratio: 4.262 Gaps: 9  
Percent Similarity: 40.653 Percent Identity: 40.315

alignment\_block:

US-09-357-675c-25 x AL354714\_5 ..  
Align seg 1/1 to: AL354714\_5 from: 1 to: 110000

1 ProLeuAlaAlaAla\*\*LeuAlaProAspArgProAspArgThrIle 17  
|||||  
3174 CCACTGCGCGCGGCTTCTGCTCCAGACCGCCCTCCGATCGGACCGT 3223  
17 uArgMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 34  
|||||  
3224 GCGAATGGTTTGGCTATATCTTCATGTAGACCTACTCCCTATCCCGTC 3273  
34 rPro..... 35  
|||||  
3274 GCGCGGGTGAATCCACCTGCGGTCTTAACTGTGAACAGAGATGC 3323  
35 ..... 35

3324 TGCCTCTGGAGAGCGGGGAGGAGCGGACCGCACTGAGCGGGCGGG 3373  
35 ..... 35  
3374 AGGGTGGAGGGCGGGCGCGCTTGGGGCCGTGGGTTCTTTGGCCCTGC 3423  
35 ..... 35  
3424 CCACAGAGGAGGGGTGGAGACGAGAGAGGTTGACTTCCCTGGCAGA 3473  
35 ..... 35  
3474 TTCTGTGAAGGGAATATGCTTCACTCACTAAGCTGCGCAAGTGC 3523  
35 ..... 35  
3524 ACAGTCAGAGAGAGACTTGGGAAACCAAGATAGTTCCCGAGATGA 3573  
35 ..... 35  
3574 CTTTGGACTGCGGAAACGTTTGTCAAGAGAAAGAGCTTCAATTAAAG 3623  
35 ..... 35  
3624 TGAGATCAATTGGAAGTTGAATATTCAGGGCGGAGACAGCCCTTCT 3673  
35 ..... 35  
3674 CCATCTCCACATGATAGTGTGTCTGTCCCTTAGCCTAATTTCTTAA 3723  
35 ..... 35  
3724 GAGCGTTTGAAGACTAGGTTTACTGCGAGGCGCTGATTCAGGCTCA 3773  
35 ..... 35  
3774 CAGTATTAAGAGAGAACTCAGAGATCTAGAAATTTATGCTTAGAGAA 3823  
36 ..... 36  
3824 GGGGTGTGTCTCTCATATCTCACTTCTCAGGCTGGGCTTCATACCA 3873  
42 rProProHisArgPheLeuSerLeuLeuCysProGlyLeuArgIlePro 58  
|||||  
3874 GGCCTCTCAGAGATTCCTGCTCTGTGTGTCGATCCGATACCT 3923  
59 GlnLeuSerValLeuCysAlaGlnPro..... 67  
|||||  
3924 CACTCTCAGTACTTGTGTCTCAGCCAGGTACACGTTTGTGTGTCC 3973  
67 ..... 67  
3974 TCAGTGCCTGGCACTTAGATGCTCACTTGTTAATGATAGTGGGAGAC 4023  
67 ..... 67  
4024 ACAGAGTGTCAACTATCCACACATTTGATTGTGAGCCCTACTAGCCCT 4073  
67 ..... 67  
4074 GGGTCAAGCTGCCCTGTAAGACATGATCAAAAGAAAGTCCAGCTTCT 4123  
67 ..... 67  
4124 GCCTTCCACTTGACCCCTTAGCATTAATTTGCTTCCCTGTGCTATGA 4173  
68 ..... 68  
4174 AATCTGAGATCTGCTATGCTGTTCACAGCGCCAGACCATGCTATC 4223  
75 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSe 91  
|||||  
4224 TCTCTTCTCTCTGGGAACTGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT 4273



CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone  
Requests: clonerequests@sanger.ac.uk  
On May 7, 2001 this sequence version replaced gi:13751001.

## COMMENT

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk

Project Information

Center project name: ba137A12

## Summary Statistics

Assembly program: XGAP; version 4.5  
Sequencing vector: Plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 164729 bases at least Q40  
Consensus quality: 165486 bases at least Q30  
Consensus quality: 166169 bases at least Q20  
Insert size: 167063; sum-of-contigs  
Insert size: 171878; 3.4% error; agarose-fp  
Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality  
coverage: 8.11x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 9373: contig of 9373 bp in length  
\* 9374 9473: gap of 100 bp  
\* 9474 29891: contig of 20418 bp in length  
\* 29892 29991: gap of 100 bp  
\* 29992 36239: contig of 6248 bp in length  
\* 36240 36339: gap of 100 bp  
\* 36340 103784: contig of 67445 bp in length  
\* 103785 103884: gap of 100 bp  
\* 103885 114878: contig of 10994 bp in length  
\* 114879 114978: gap of 100 bp  
\* 114979 126347: contig of 11369 bp in length  
\* 126348 126447: gap of 100 bp  
\* 126448 137764: contig of 11317 bp in length  
\* 137765 137864: gap of 100 bp  
\* 137865 163160: contig of 25296 bp in length  
\* 163161 163260: gap of 100 bp  
\* 163261 167863: contig of 4603 bp in length.

## FEATURES

## source

1. 167863

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/clone="RP11-137A12"

/clone\_id="RP11-11.1"

1. 9373

/note="assembly\_fragment:02360"

fragment\_chain:1

clone\_end:"7"

vector\_side:left"

9474. 29891

/note="assembly\_fragment:01854"

fragment\_chain:1"

29992. 36239

/note="assembly\_fragment:00805"

fragment\_chain:1"

36340. 103784

/note="assembly\_fragment:01467"

fragment\_chain:1"

103885. 114878

/note="assembly\_fragment:00991"

fragment\_chain:2"

114979. 126347

/note="assembly\_fragment:01503"

misc\_feature  
126448. 137764  
/note="assembly\_fragment:01297"

fragment\_chain:2"

misc\_feature  
137865. 163160  
/note="assembly\_fragment:00117"

fragment\_chain:2"

misc\_feature  
163261. 167863  
/note="assembly\_fragment:01267"

fragment\_chain:2

clone\_end:SP6

vector\_side:right"

BASE COUNT 43927 a 39935 c 40002 g 43197 t 802 others

ORIGIN

## alignment\_scores:

Quality: 1538.50 Length: 888  
Ratio: 4.262 Gaps: 9  
Percent Similarity: 40.653 Percent Identity: 40.315

## alignment\_block:

US-09-357-675c-25 x AL590651/rev ..

Align seg 1/1 to reverse of: AL590651 from: 1 to: 167863

1 ProLeuAlaIAlaIa\*\*LeuAlaProAspArgProProAspArgPhe 17

|||||

152076 CCACCTGCTCGGCGCTTCTGCTCCAGACCGCCCTCCGATGGACCT 152027

17 uArgMetValLeuAlaIleSerSerCysArgThrTySerLeuSerArg 34

|||||

152026 GCGAATGGTTTGGCTATATCTTCATGTAGACCTACTCCATGCCGTC 151977

34 rArgPro..... 35

151976 GGCGGGGGTGAATCCACCTGCGTCTTAACTGTGTACAGAGATGC 151927

35 ..... 35

151926 TGCCTTGGAGAGCGCGGAGGAGCGGCAACTGAGCGGGCGGCGG 151877

35 ..... 35

151876 AGGGTGAGGCGGCGGCGGCTTGGGCGCTGCTTCCCTTGCCTTGC 151827

35 ..... 35

151826 CCACCAAGGAGGGGTGGAGAGAGAGGTGAACCTTCCCTGCGAGA 151777

35 ..... 35

151776 TTCTGGTGAAGGGAATATGCTTGCAGTCAAGTAAAGCTGCCAAGTGC 151727

35 ..... 35

151726 ACAGTCAAGAGAGAGCTTGGGAAAAACAAGATAGTCCCGAGATGA 151677

35 ..... 35

151676 CTTTGGACTGCGAAACGTTGTCAAGAGAAAGAGCTTCAAGTTAANG 151627

35 ..... 35

151626 TGAGATCATTTGGAAGTTGAACATATTCAGCGGCGGAGAGAGCCCTTTC 151577

35 ..... 35

151576 CCATCTTCCACTGTATAGTCTGTCCCTTAGCCTAATTTCTTATAAA 151527

35 ..... 35

151526 GAGCGTTTGAAGACTAGATTTTACTGCGAGGCGCCGTGATTCAGGCTCA 151477

```

35 ..... 35
151476 CAGTATTAAGAGAGAACTCAGAGATCCTAGAATTATTCCTTGAGAA 151427
36 ..... ArgLeuGlyPheIleThra 42
151426 GGGGTTGGTTCCTCATATCTCACCTTCCTCAGAGCTGGGCTTCATCACC 151377
42 rPrProhHisArgPheLeuSerLeuLeuCysProGlyLeuAlaGlyIlePro 58
151376 GGGCTCCCTCAGACATTCCTGTCCTTCGTCTGACACCCGAAACCT 151327
59 GlnLeuSerValLeuCysAlaGlnPro ..... 67
151326 CAACCTCTCAGTACTTGTCTCAGCCAGGTACACGCTTTGTGTGTCC 151277
67 ..... 67
151276 TCAGTGCCTGGCAGCTTAGATGCTCAGTTGTTAAATGAGATGGGAGAC 151227
67 ..... 67
151226 ACAGAGTGTCAACTATCCACACATTTGATGTGAGCCCTACTAGCCCT 151177
67 ..... 67
151176 GGGTCACAGTGCCTGTAAAGACATATCAAAAGAAATCCAGCTTTCCT 151127
67 ..... 67
151126 GCCTTCACATTGCACCCCTTAGCATTAATTTGCTTCCTGTGCTATGA 151077
68 ..... ArgProAlaMetAlaIle 74
151076 AATCTGAGATCCTGCCTATGCTGTTCACAGGCCAGAGCCATGCTATC 151027
75 SerSerSerCysGlnLeuProLeuValAlaValCysGlnValIhrse 91
151026 TCCTTCTCCCTCGGACACTGCCCTGGTGCTGTGTGCGAGGTAACATC 150977
91 rThrProAspLysGlnGlnAsnPheLysThrCysAlaGluLeuValArg 108
150976 GAGCCAGACAAACACAGAACTTTAAACATGCTGCTGACCTGTTCCAG 150927
108 LysAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaIhrAsp 124
150926 AGGCTCCAGACTGGGCTGCTGCTGCTTCTGCTGAGGCAATTTGAC 150877
125 PheIleAlaArgAspProAlaGlnThrLeuHisLeuSerGluProLeuG 141
150876 TTCATTGACAGGGACCTGCAGAGAGCTACACCTGCTTAACACACTGG 150827
141 yGlyLysLeuLeuGluGlnIleThrGlnLeuAla ..... 152
150826 TGGGAACCTTTTGAAGAATACACCCAGCTTGCAGAGTATCAGGAATA 150777
152 ..... 152
150776 GCGAGGAGAGAGTAGAATCTTTGTGACAGTGTCCCTGGTTGCCAGAT 150727
152 ..... 152
150726 ATGAGGGTAGAGCCTTGAGAACTCAGTGAAGAGTTGCTCAGTGCCTTCC 150677
153 .... ArgGluCysGlyLeuThrPheLeuSerLeuGlyGlyPheHisGluArg 168
150676 CCCAGAGGAATGTGACTGTGGCTGTCTGGGTGTTTCCATGAGGCTG 150627
168 LysAlaSerTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeu 184
150626 GCCAAGACTGGAGCAGAGACTCAGAAAATCTAACAAATGTCAAGTGTCTG 150577

```

```

185 AsnSer..Lys ..... 187
150576 AACAGCAAAAGGTAGACTTTATAACCTTAGCCTGTGCTTCCCATGCT 150527
187 ..... 187
150526 CTTCACCTAGATTCTCCAGAAATTGTTTCAACTCTATTATTCCTTGACC 150477
187 ..... 187
150476 CAAGATTAAAGGGGTGCTCTACTACTAGTTCCTAGCCCTATTAACATATCTC 150427
187 ..... 187
150426 CTCCTTGGGAGAGTAAGCAAGCCTCTAGAACACACGACATATATCC 150377
188 ..... GlyAlaValAlaIleThrTyrArgLysThrHisLe 199
150376 TTCTTCTTACTGTAGGGCAGTAGTGCCACTTACAGAAACACATCT 150327
199 uCysAspValGluIleProGlyGlnGlyPrometCysGluSerAsnSer 216
150326 GTGTGACGTAGAGATTCCAGGGCAGGCTTATGTGAAAGCAACTCTA 150277
216 hMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLys 232
150276 CCATGCTGGGCCAGCTTGTGAGTCACTGTACAGCACACGACGAGCAAG 150227
233 IleGly..Leu ..... 235
150226 GTAGAGATTGTAAAGAGATGAGGAGGGAACAGAAATCTTGAATCG 150177
235 ..... 235
150176 CAGTAGAGATAGAAACCCCTAAGAGAGGGGTAAATGAATAATGACTAG 150127
235 ..... 235
150126 ATGCTGTGACAAACAGACGAGGAAGACTACTAAGTAGGCTTTTTCAT 150077
236 ..... AlaValCysTyrAspMetLargPheProGluLeuSerL 248
150076 CCAGATTGCTGTAGCTGTGTATGACATGCGGTTCCCGAATCTCTC 150027
248 euAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 264
150026 TGGCATTGGCTCAAGCTGGAGAGATGACTTACTTCTTCCCTTACGCTTT 149977
265 GlySerIleThrGlyProAlaHisTrpGluVal..Leu ..... 276
149976 GGATCCATTACAGGCCCAAGCCACTGGGAGGTAAAGATGATGCCTTTAA 149927
276 ..... 276
149926 AACATAGGGCCTTTTCTTAACCTCATCTTCCCCCTTGGGCCATACAGT 149877
276 ..... 276
149876 TAAATTCCTTCCCTTCCACCTAATGGAAACTCATTTCCCGAGATAT 149827
276 ..... 276
149826 TCTCTCATGATAGTTAAATAGTAATCATTTCTTAGGCAATTAACAAA 149777
276 ..... 276
149776 TAGTCACAATGGGTAGATGTGTGTATGATGCCCTGACCTGTACTTCC 149727
276 ..... 276
149726 CACTATTGTCTACATGACTTAAGTAGAACACATCTCATGCCCAGGTGT 149677
277 ..LeuArgAlaArgAlaIleGlnThrGlnCysTyrValValAlaAlaAla 292

```

```

149676 TCGCTGGGCGCCGCTATCGAAACCCAGTGTATAGTGACGACGCA 149627
293 GINCYSGIARIGHISIGLULYSARGALASERTYGLYHISSEMETYA 309
149626 CAGTGTGGACGCGACCATGAGAGCAAGTTATGGCCACGCAWTGGT 149577
309 IVALASPPRTPRGlyThrValValAlaIaArgCYSserGIUGlyProGlyL 326
149576 GCTAGACCCCTGGGAGACAGTGTGGCCGCTGCTGAGGGCCAGGCC 149527
326 euCYsLeuAlaArgIleAspLeuAsnTYrLeuArgGlnLeuArgArgHis 342
149526 TCTGGCTTCGCCAATAGACCTCAACTATCTGCGAGCTTGGCCGACAC 149477
343 LeuProValPheGlnHisArgArgProAspLeuTYrGlyAsnLeuGlyH 359
149476 CCGCTGTTCACGACCGCGGCTGACCTATAGGCAATCTGGGTCA 149427
359 sProLeuSer 362
149426 CCCACTGTCT 149417
seq_name: gb_pr:AL591806
seq_documentation_block:
LOCUS AL591806 200822 bp DNA linear PRI 30-JAN-2002
DEFINITION Human DNA sequence from clone RP11-544M22 on chromosome 1, complete
sequence.
ACCESSION AL591806
VERSION AL591806.16 GI:18476709
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 200822)
AUTHORS Harrison, E.
TITLE Direct Submission
JOURNAL Direct Submission
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail requests: clonesrequest@sanger.ac.uk
humquertys@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:117902927.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Emu, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/chr1
RP11-544M22 is from the library RPI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/Dacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-544M22. The true
left end of clone RP11-137A12 is at 156538 in this sequence. The
true right end of clone RP11-381D2 is at 145015 in this sequence.
Location/Qualifiers

```

```

source
1. .200822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-544M22"
/clone_1lb="RPI-11.2"
10395. .10397
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
44098. .44107
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
59585. .59604
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
90797. .90883
/note="Single clone region. Reads generated from a
transposon library derived from a single pUC clone.
Restriction digest data confirm the assembly."
90861. .90883
/note="Sequence from uni-directional dgrp big dye
terminator reads only."
BASE COUNT 55215 a 49052 c 46183 g 50372 t
ORIGIN
alignment_scores:
Quality: 1538.50 Length: 888
Ratio: 4.262 Gaps: 9
Percent Similarity: 40.653 Percent Identity: 40.315
alignment_block:
US-09-357-675c-25 x AL591806 ..
Align seg 1/1 to: AL591806 from: 1 to: 200822
1 ProLeuAlaAlaAla**LeuAlaProAspArgProAspArgThrLe 17
172105 CCACCTGCTGCGCGGCGCTCTCCAGACCGGCTCCGAGTCGACCT 172154
17 UArgMetValIleuAlaIleSerSerCysArgThrTYrSerLeuSerArg 34
172155 GCGAAAGGTTTGGCTATCTTCATGAGGACCTCCATCCGCGTC 172204
34 rgPro.....
172205 GCGCGGCTGAATCCCACTGCGGCTTAACTTGTACAGAGATGC 172254
35 .....
172255 TGCCTGTGGAGAGCGGGAGGAGGACGCAACTGAGCGGGCGGG 172304
35 .....
172305 AGGCTGAGAGCGGGCGGCGCTTGGGCGCTTGCTTCCCTTGC 172354
35 .....
172355 CCACGAGGAGGGGTGGAGACGAGAGAGGTGAATTCCTCCCTGGAGA 172404
35 .....
172405 TTCTGTGAAGGAGGAATATGCTTCGAGTCAGTAAGCTCGCAAGTGC 172454
35 .....
172455 ACAGTCACAGAGAGAGCTTGGGAAACCAAGATATTCGCGAGATGA 172504
35 .....
172505 CTTTGGACTGGGAAACGTTGTCAAGAGAAAGAGCTTCACTTAATG 172554
35 .....
Location/Qualifiers

```

172555 TGAGATCATTTGAAGTGAATCTATTCAGCGGCGAGAGAGCCCTTCT 172604  
 35 ..... 35  
 172605 CCATCTTCCACTGTATGCTGTGCTCCCTAGCCTAATTTCTTATAAA 172654  
 35 ..... 35  
 172655 GAGGCTTTGAGACTAGGTTTTTACTGCGAGGGCCCTGATTCAGGCTCA 172704  
 35 ..... 35  
 172705 CAGTATAAGAGAGAGAGTCAAGAAATCTAGAAATTTATTGCTTGAAGAA 172754  
 36 ..... ArgLeuGlyPheIleThr 42  
 172755 GGGGTGGTGTCTCTCATATCTCACTTCTCAGGCTGGCTTCATTCACCA 172804  
 42 rfpProH1sArgPheLeuSerLeuLeuGlyLeuArgIlePro 58  
 172805 GGCCTCTCCACAGATTCTCTCCCTCTGTGTCTGACTCCGATACCT 172854  
 59 GlnLeuSerValLeuCysAlaGlnPro ..... 67  
 172855 CAACTCTCAGTACTTGTGCTCAGCCAGGTAACAGCTTTTGTGTGTCC 172904  
 67 ..... 67  
 172905 TCAGTGCCCTGGCACTTAGATGCTCAGTTGTTGTAATGATAGTGGAGAC 172954  
 67 ..... 67  
 172955 ACAGAGGTGCACACTATCCACACTTTGATTTGGTACCCCTACTAGCCCT 173004  
 67 ..... 67  
 173005 GGGTCACAGTGGCCCTGTAAAGAGCATGATCAAAAGAGTCCAGTTCT 173054  
 67 ..... 67  
 173055 GCCTCTCCACTTGCACCCCTTAGCATTAATTTGCTTCCCTGCTATGA 173104  
 68 ..... ArgProArgAlaMetAlaIle 74  
 173105 AATCTGGAATCTGCTTATGCTGTTCACAGGCCAGAGCCATGCTATC 173154  
 75 SerSerSerSerCysGluLeuProLeuValAlaValCysGlnValIlePro 91  
 173155 TCCGTCTCTCCGCGAAGTGGCCCTGTGCTGTGTGCCAGGTACATC 173204  
 91 rTrpProAspLysGlnGlnAsnPhelyTrpCysAlaGluLeuValArg 108  
 173205 GAGGCCAGACAGCAAGCACTTAAACATGTGCTGAGCTGGTTCGAG 173254  
 108 IuAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheasp 124  
 173255 AGGCTGCCAAGCTGGGTGCTGCTGCTTCTGCTGAGGCAATTGAC 173304  
 125 PheIleAlaArgAspProAlaGluTrpLeuHisLeuSerGluProLeuG 141  
 173305 TTCATTGCAGGGAGCCCTGCAGAGAGCTACACCTGTCTGACACAGTGG 173354  
 141 YGlyLysLeuLeuGluGluTrpTrpGlnLeuAla ..... 152  
 173355 TGGGAAACTTTTGGAAATACACCCAGCTTGCAGGATATCAGGAATA 173404  
 152 ..... 152  
 173405 GCGAGGAGAGTAGAATCTTTGTGACAGTGTCCCTGGGTGCCAGAT 173454  
 152 ..... 152

173455 ATGAGGCTAGAGCCCTTGAGAGTCAAGTGAAGAGTTGTAGTGTCCCTCC 173504  
 153 ... ArgGluCysGlyLeuTrpLeuSerLeuGlyPheHisGlnArg 168  
 173505 CCCAGGGAATGTGACTGTGGCTGTCTGTGGGTGGTTTCCATAGAGGTG 173554  
 168 LysGlnAspTrpGluGlnTrpGlnLysIleTrpAsnCysHisValLeu 184  
 173555 GCCAAGACTGGGACAGACTCAGAAATCTACATTTGACAGTGTGCTGTG 173604  
 185 AsnSer...Lys ..... 187  
 173605 AACGCAAGGTGAGACTTTATACCTTTAGCCTGCTCTCCATGCT 173654  
 187 ..... 187  
 173655 CTTCACCTAGATTCTCCAGAAATGTTCTCAGCTTATTTCTTGACC 173704  
 187 ..... 187  
 173705 CAAGATTAGGGGTGCTCTACTTCACTTCTAGCCTATTAACATATCTC 173754  
 187 ..... 187  
 173755 CTCCTGGAGAGAGTAAGCAAGGCTTCTAGAAACACAGCACTGATATTC 173804  
 188 ..... GlyAlaValAlaIleThrTrpArgLysIle 199  
 173805 TTCCTTTCTACTGTAGGGGAGTGTGCTCCTTACAGAGGAGACACATCT 173854  
 199 uCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 216  
 173855 GTTGACGCTAGAGATTCCAGGGCAGGGGCTTATGTGCAAGCAACTCTA 173904  
 216 hMetProGlyProSerLeuGluSerProValSerTrpProAlaGlyLys 232  
 173905 CCATGCCGTGGGCCAGTCTTGATGACCTGTGACGACACACGAGGCAAG 173954  
 233 IleGly...Leu ..... 235  
 173955 GTAGAGTTGTGAAGATGAGAGGAGGGGAAACAGAAATACTTTGAACGTG 174004  
 235 ..... 235  
 174005 CAGTAGAGATAGAAAGCCCTAAGAGAGGGGTAATGAAATATGACTAG 174054  
 235 ..... 235  
 174055 ATGGTGTGACAAACAGAGCAGAGAGACTACTAAGTAGGCTGTTTTTCATT 174104  
 236 ..... AlaValCysTrpAspMetArgPheProGluLeuSerL 248  
 174105 CCAGATTGCTTACCTGTCTGTATGACATCGGTCCCTCAACCTCTCTC 174154  
 248 euAlaLeuAlaGlnAlaGlyAlaGluIleLeuTrpTrpSerAlaPhe 264  
 174155 TGGCATTTGGCTCAGAGCTGGAGCAGAGATACTTACCTTACCTTCAGCTTT 174204  
 265 GlySerIleThrGlyProAlaHisTrpGluVal...Leu ..... 276  
 174205 GGATTCATTAAGGCCAGCCGACCTGGGAGGTAAAGATATGCTTTTAA 174254  
 276 ..... 276  
 174255 AACATAAGGCTTTTCTTAACCTCATCTCCCTGGCCATCAGAGT 174304  
 276 ..... 276  
 174305 TAAATTCCTTCCCTTTCACACTAATGGGAAACTCATTTCCCCAGATATT 174354  
 276 ..... 276  
 174355 TCTCTCATGATAGTTAAATAGTAAATCATTTCTAGGCAATTAACAAA 174404

```

276 ..... 276
174405 TACTCACATGGGTAGATGGTGTGTAATGTCCTCAGCTGACTCTCC 174454
276 ..... 276
174455 CACTATTGCTACATGACTTAAGTGACACACATCTCATGCCAGGTGT 174504
277 .. LeuArGAlaArGAlaIleGluThrGlnCysTyrValValAlaAla 292
174505 TCGTGGGGGCCCGTGTATGAAACCCAGTGTATGAGGACAGCA 174554
293 GlnCysGlyArgHisHisGlnLysArgAlaSerTyrGlnHisSerMetVa 309
174555 CACTGTGGAGCCACCATGAGAGAGCAAGTTATGCCACAGCATGTT 174604
309 ValAspProTyrPglyThrValValAlaArgCysSerGlnGlyProGlyL 326
174605 GGTAGACCCCTGGGGAACTGGTGGCCCTGCTGTGAGGGCCAGGCC 174654
326 euCysLeuAlaArgIleAspLeuAsnTyrLeuArGlnLeuArGArgHis 342
174655 TCTGCTTGGCCGAATGACCTCACTATCTGGACAGTTGGCGCGCAC 174704
343 LeuProValPheGlnHisArgArgProAspLeuTyrGlnAsnLeuGlyH 359
174705 CTGCGGTGTTCAGACGCCGAGCCCTGACCTCATGTGCATCGGTCA 174754
359 sProLeuSer 362
174755 CCCACTGCTT 174764
seq_name: gb_pr:AF069984
seq_documentation_block:
LOCUS AF069984 4079 bp DNA linear PRI 23-JUL-1998
DEFINITION Homo sapiens nitrilase homolog 1 (NIT1) gene, alternatively spliced
product, complete cds.
ACCESSION AF069984
VERSION AF069984.1 GI:3242977
KEYWORDS
SOURCE
ORGANISM human.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 4079)
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrilase and fhl homologs are encoded as fusion proteins in
drosophila melanogaster and caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
JOURNAL 98337986
MEDLINE 2 (bases 1 to 4079)
Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES
FEATURES
SOURCE
1..4079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q21-q22"
join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3569)
/gene="NIT1"
/note="alternatively spliced"
/product="nitrilase homolog 1"
635..3569
gene

```

```

exon
635..680
/gene="NIT1"
/note="alternatively spliced"
CDS
1016(679..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3569)
/gene="NIT1"
/note="NIT1; alternatively spliced"
/codon_start=1
/product="nitrilase homolog 1"
/protein_id="AAC39901.1"
/db_xref="GI:3242978"
/translacion="MLGFITRPFRFLSLPCGLRIPOLSVLCAQPRPRMAISSSSC
ELPLVAVQVSTPDKOONFKTCAELVEARLGLAFLPEAFDFIARDAETLHS
EPKGLLEETVTLAREGIMLSGGFHERODMEOIKYCNHVLNASKAVATYR
KTHGLDVEIRPGGPMPCESNSTMPGSLSPSTPRAGKTGLAVCYDMRPPSLALAAQ
GAELITVPSAFSGSITGPAHWEVLLKRAIETQCYVAAAGGRHHEKRASTGHSWVD
FWGITVVARCSSEPGICLARIDLVRLRLPVRQHRPDLXGNLGHPLS"
681..711
/gene="NIT1"
/note="1A; alternatively spliced"
712..741
/gene="NIT1"
/note="1B; alternatively spliced"
940..1053
/gene="NIT1"
/note="1C; alternatively spliced"
1288..1383
/gene="NIT1"
/number=2
/number=2
1636..1890
/gene="NIT1"
/number=3
2011..2114
/gene="NIT1"
/number=4
2321..2454
/gene="NIT1"
/number=5
2609..2734
/gene="NIT1"
/number=6
3001..3569
/gene="NIT1"
/number=7
exon
BASE COUNT 940 a 1082 c 1043 g 1012 t 2 others
ORIGIN
alignment_scores:
Quality: 1525.50 Length: 889
Ratio: 4.226 Gaps: 10
Percent Similarity: 40.607 Percent Identity: 40.270
alignment_block:
US-09-357-675C-25 x AF069984
Align seg 1/1 to: AF069984 from: 1 to: 4079
1 ProLeuAlaAlaAla**LeuAlaProAspArgProProAspArgThrIle 17
||||| 17
604 CCACGCTGCTGGCGGCTTGTGCTCCAGACCGCCCTCGGATGGAGACCT 653
17 uArg.MetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 33
||||| 33
654 GCCAAATGTTTGGCTATATCTTCAATGAGACCTACTCCATCCCGT 703
34 ArgPro..... 35
704 CGGCGCGGTGAATCCACCTCGGCTTTAACTGTGTGAACAGAGATG 753
35 ..... 35

```



```

754 CTGCCTGTGGAGAGCGGGGAGGAGCAACTGGAGCGGGCGG 803
35 ..... 35
804 GAGGGTGAAGGGCGGGCGGCTTGGGGCTGGTTCCTTCCCTTG 853
35 ..... 35
854 CCCACAGGAGGGGTGGGAGACGAGAGGGTGAACCTTCCCTGCGAG 903
35 ..... 35
904 ATTCTGTGAAAGGGGAAATATGCTTCGAGTCAGTAAAGCTGGCAAGTG 953
35 ..... 35
954 CACAGTCAAGAGAGAGCTTGGGAAACCAAGATATGTTCCCGAGATG 1003
35 ..... 35
1004 ACTTTTGACGTGGGAAACGTTTGTCAAGAGAAAGGCTTCAGTTTAT 1053
35 ..... 35
1054 GTGAGATCATTTGAAGTTGAACATATTCAGGGCGGAGCAGAGCCCTTTC 1103
35 ..... 35
1104 TCCATCTCCCACTGATATGCTGTTCCTTACCTATATATTTCTTATAA 1153
35 ..... 35
1154 AGAGCGTTTGAAGACTAGGTTTACTGCGAGGCCCTGATTCAGGCTC 1203
35 ..... 35
1204 ACAGTATAAGAGAGAGTCAGAGAACTCAGAAATTTATGCTTGAGA 1253
36 .....
1254 AGGGGTGTGCTCCTATATCTCAGCTTCTCAGGCTGGGCTTCATCACC 1303
42 ArgProProlHisArgPheLeuSerLeuCysProGlyLeuArgIleP 58
|||||
1304 AGGCTCTCTCAGATTCCTGCTCTGTCTGTCTGTGACTCCGGATACC 1353
58 ogInLeuSerValLeuCysAlaInPro..... 67
|||||
1354 TCAACTCTCAGTACTTGTGTCTCAGCCAGGTAAACGTTTGTGTGTC 1403
67 ..... 67
1404 CTCAGTGCTGACACTTAGATGCTCAGTTTGTAAATGATAGTGGGAGA 1453
67 ..... 67
1454 CACAGAGATGTCACATTCACACATTTGATGGTGAGCCCTACTAGCCC 1503
67 ..... 67
1504 TGGGTCAACGTGCCCTGTAGAGCATGATCAAAAGAGATCCAGCTTCC 1553
67 ..... 67
1554 TGCCCTTCACACTGCACCCCTAGCATTTAAATTTGCTTCCCTGTATG 1603
68 ..... ArgProArgIleMetAlaIle 74
|||||
1604 AAATCTGAGATCCTGCTATGCTGTTCACAGGCCAGAGCCATGGCTAT 1653
74 eSerSerSerCysGluLeuProLeuValAlaValCysGluValThrS 91
|||||
1654 CTCCTCTTCCCTCGGAACTGCCCTGTGCTGTGTGCCAGGTAAAT 1703
91 erThrProAspLysGlnGlnAsnPheLeuThrCysAlaGluLeuValArg 107
|||||
1704 CGAGCCGACAGCAAGACACAGAACTTTAAACATGTGTGAGTGTGCA 1753
108 GluAlaIleArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAs 124
|||||
1754 GAGGCTGCCAGACTGGGTGCTGCTGTGCTTCTGCTGAGGATTTGA 1803
124 pheIleAlaIleArgAspProAlaGluThrLeuHisLeuSerGluProLeuG 141
|||||
1804 CTTCATTTGACAGGGAGCCCTGACAGACGTTACACCTGTGAACCACTGG 1853
141 LysGlyLysLeuLeuGluGluTyrThrGlnLeuAla..... 152
|||||
1854 GTGGAAACTTTTGGAGAAATACACCAGCTTGCCAGGTATCAGGAAAT 1903
152 ..... 152
1904 AGCGAGGAGAGGTAGAAATCTTTGTGACAGTGTCCCTGGTTGCCAGA 1953
152 ..... 152
1954 TATGAGGGTAGAGCCTTGAGAACTCAGTGAAGAACTGTGCATGTCCTTC 2003
153 ..... ArgGlyCysGlyLeuThrLeuSerLeuGlyGlyPheHisGluArg 167
|||||
2004 CCCCAGAGGAATGTGAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 2053
168 GlyIleAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLe 184
|||||
2054 GGCACAGACTGGGAGCAGACTCAGAAATCTACATTTGTACGTGCTGCT 2103
184 uAsnSer..Lys..... 187
|||||
2104 GAACAGCAAAAGGTGAGACTTTATTAACCTTTAGCCTGCTGTCCATGC 2153
187 ..... 187
2154 TCCTTCTACATAGATTCACAGAAATGTTCACAACTCTATTTCTTGAC 2203
187 ..... 187
2204 CCAAGATTAGGGGTGCTCTACTTCACTTCTAGCTTAACCTATPCT 2253
187 ..... 187
2254 CTTCTTGGGAGAGGTAAAGCAAGCTTCTAGAACCAACGACCTGATATTC 2303
188 ..... GlnAlaValValAlaThrTyrArgLysThrHisL 199
|||||
2304 CTCTTCTTCTTACTGTAGGGGACATGAGTGGCCACTTACAGGAAGACATC 2353
199 eucYasPvalGluIleProGlyGlnGlyPrometCysGluSerAsnSer 215
|||||
2354 TGTGTAGCTAGAGATTCCAGGGGAGGAGGCTATGTGGAAGAACTCT 2403
216 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyL 232
|||||
2404 ACCATGGCTGGGCCAGCTTGTGAGTCACTGTCACTGTCAACACAGAGGCA 2453
232 sIleGly..Leu..... 235
|||||
2454 GTTAGGAGATTGTGAAGAGATGAGGAGGAGGAACAGAAATCTTGAATG 2503
235 ..... 235
2504 GCAGTAGAGATGAAGACCCTAAGAGAGGGGTAAATGAATATGTACTA 2553
235 ..... 235
2554 GATGCTGTACAAACAGAGAGCAAGAACTACTAGTAGGCTGTTTTCAT 2603

```

```

236 .....AlaValCysTyrAspMetArgPheProGluLeuSer 247
      |||||
2604 TCAGATTTGGTCTAGCTGTCTGTATGACATGGGTTCCCTGACACTCT 2653
      |||||
248 LeuAlaLeuAlaGlnAlaGlyAlaGluLeuThrTyrProSerAlaPh 264
      |||||
2654 CTGGCATTTGGCTCAGCTGGAGCAGAGATTAATCTATCCCTTACAGCTT 2703
      |||||
264 egiysrliethrGlyProAlaHisTrpGluVal..Leu..... 276
      |||||
2704 TGGATTCATTACAGGCCACAGCCAGCTGGAGGTAAGATGATGCCCTTTT 2753
      |||||
276 ..... 276
2754 AATCATTAAGGGCTTTTCTTAACCTCATCTTCCCCCTTGGCCCTACCA 2803
      |||||
276 ..... 276
2804 TTAAATTCCTTCCCTTCCACCTAATGGGAAACATCATTCCTCCAGAT 2853
      |||||
276 ..... 276
2854 TTCTCTCAGTAATAGTTAAATAGTAATCATCTCTAGGCAATTACCAA 2903
      |||||
276 ..... 276
2904 ATAGTCACATGGGATGATTGCTGTAAATGTCCTCACCTGTCACCTTCC 2953
      |||||
276 ..... 276
2954 CCACATTTCTCATGTAATGTAAGTGAACACACATCTCATGCCAGGTG 3003
      |||||
277 ...LeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAla 292
      |||||
3004 TTGCTCGGGGCCGCTGCTATCGAACCAGCTATGCTAGTGGCAGCAGC 3053
      |||||
292 aglncysgllyargHisIsglulysargAlaSerTyrGlyHisSerMet 309
      |||||
3054 ACAGTGTGGAGCGCCACCATGAGAGAGAGCAATTTATGGCCACAGCATG 3103
      |||||
309 aValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGly 315
      |||||
3104 TGGTACAGCCCTGGGAGACAGTGTGGCCGCTGCTGAGGGGCGCAGGC 3153
      |||||
326 LeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgH 342
      |||||
3154 CTCTGCTTGGCCGAAATAGACTCTCACTATCTCGACAGTTGGCGCAGCA 3203
      |||||
342 sLeuProValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGly 359
      |||||
3204 CCGGCTGTGTTCCAGCAGCGAGGCTGACCTATGAGCAATCTGGGTCT 3253
      |||||
359 IsProLeuSer 362
      |||||
3254 ACCCACTGTCT 3264
      |||||
seq_name: gb_to:AF069988
seq_documentation_block:
LOCUS AF069988 1338 bp mRNA linear ROD 23-JUL-1998
DEFINITION Mus musculus nitrilase 1 (Nlt1) mRNA, complete cds.
ACCESSION AF069988
VERSION AF069988.1 GI:3228667
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1338)
REFERENCE
1 Bekarsky,Y., Campigillo,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

```

```

TITLE
Nitrilase and fhl1 homologs are encoded as fusion proteins in
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
MEDLINE
98337986
REFERENCE
2 (bases 1 to 1338)
AUTHORS
Bekarsky,Y., Campigillo,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
TITLE
Submitted Submission
JOURNAL
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES
source
1..1338
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/map="1q21-q23"
1..1338
/gene="Nlt1"
58..1029
/gene="Nlt1"
/codon_start=1
/product="nitrilase 1"
/protein_id="AAC40185.1"
/db_xref="GI:3228668"
/translation="MLGFTTRPPHQLCTGYRLTRTPVLTQPPRRMTSSSTSMELPL
VAVCOVSTPNKQENFTCAELVOEARLFLAPLPAFPIARNPETLLSEPLN
GDILGOVSALARBEGLIMSLGFERGDMONKIKYCHVLSKSGSVASVYRTHL
CDVEIPROGPRRESNYTRPGSTLEPPYKTPGKGLACAYDMRPELSIKLAQCAET
LTPSARFSVTPAHMEYLARAIIESQVIAAAGCCRHETASGHSNVDPWGT
VVARCSBPGICLARIDHFLQDMRHLVPRHRRPDLXSLGHPLS"
BASE COUNT 347 a 335 c 330 g 326 t
ORIGIN
alignment_scores:
Quality: 1496.00 Length: 356
Ratio: 4.704 Gaps: 3
Percent Similarity: 89.326 Percent Identity: 80.337
alignment_block:
US-09-357-675c-25 x AF069988
Align seg 1/1 to: AF069988 from: 1 to: 1338
7 LeuAlaProAspArgProProAspArgThrLeuArgMetValLeuAlaI 23
|||||
2 CTGGCTCTGGCAGCTCTAGAGATCCGACCCGCTGATGTTTGGCTGT 51
|||||
23 eSerSerCysArgThrTyrSerLeuSerArgArgProArgLeuGlyPhe 40
|||||
52 GTCTTCATGCTG.....GGCTTCA 70
|||||
40 LeThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLeuArg 56
|||||
71 TCACCAAGGCTCTTCCACCAA.....CTCCGTGTACCGGATCCGA 111
|||||
57 IleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAl 73
|||||
112 TTACTTGCAGACCCAGTACTTGTACTACAGCCAGGCCAGAACCATGTC 161
|||||
73 aIleSerSerSerSerCysGluLeuProLeuValAlaValCysGlnVal 90
|||||
162 C...TCATCACTTCTCTGGAGAGCTGCCCTGGTGGCTGTGTCAGAGTAA 208
|||||
90 hsrThrProAspLysGlnGlnAsnPhelYsThrCysAlaGluLeuVal 106
|||||
209 CATCAACACCAACAAGCAAGCAAGCAATTTAAACATGTGCTGATGGTGT 258
|||||
107 ArgGlnAlaAlaArgLeuGlnGlyAlaCysIleuAlaPheLeuProGlnAlaPh 123
|||||
259 CAAGAGGCTGCCAGACTGGTGTCTTCTGCGCTTCTGCTGAGGCATT 308
|||||

```

```

123 eAspPheILeAlaArgSPRoAlaGluThrLeuHisLeuSerLuprol 140
|||||
309 TGACTTATATGACAGAAACCTGGCGAGACTATCTCTGCGAACCAC 358
|||||
140 euGlyLyLyLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGly 156
|||||
359 TGAATGGGGATCTTTGGCCAAATATAGCCAGCTTCCAGGAGATGTGGA 408
|||||
157 LeuThrLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTyrGlu 173
|||||
409 ATCTGGCTGCTTGGCGGTTTCCAGAGCGTGGCCAAAGCTGGAGCA 458
|||||
173 nThGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAla 190
|||||
459 GAATCAGAAAATCTACAAATCTCATGTGCTTTTGAACAGCAGGATCAG 508
|||||
190 aValAlaThrTyrArgLysThrHisLeuCysAspValGluIleProGly 206
|||||
509 TAGTGGCCAGTTACAGAGACACATCTGTCCGATGATGATCCAGGT 558
|||||
207 GlnGlyProMetCysGluSerAsnSerThrMetProGlyProSerLeuG 223
|||||
559 CAGGGCCGATGAGAGAAAGCACTATACCAAGCTGGAGGACCTTTGA 608
|||||
223 UserProValSerThrProAlaGlyLysIleGlyLeuAlaValCysTyr 240
|||||
609 GCCACCTCTCAAGACACCGCTGGCAAGTTGGTCTGCAATCTGTTATG 658
|||||
240 sPeKtArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 256
|||||
659 ACATGGCGTTCCTGCACTTCTTTGAAATTTGGCTCAAGCTGGGGCAGA 708
|||||
257 IleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHis 273
|||||
709 ATACTTCTATCCCTTCAGCTTGTGATCTGTACAGTCCAGCCACTGT 758
|||||
273 pGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValAla 290
|||||
759 GGAGTCTCTGCTGGGGCCCGCCATTTGAATCTCAGTGCATTAATAG 808
|||||
290 IaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyHis 306
|||||
809 CAGCAGCGCAGTGTGAGCGCACCATGAACACAGCAAGTATGGCCAT 858
|||||
307 SerMetValValAspProTyrGlyThrValAlaAlaArgCysSerGlu 323
|||||
859 AGCATGGTGTGACCGTGGGCGACAGTGGTGGCCGCTCTCCGAGGG 908
|||||
323 YProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeu 340
|||||
909 ACCAGGCTCTGCTTGCATGATGATCTTCACATTCTACACAGATGC 958
|||||
340 rGArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGlyAsn 356
|||||
959 GCCAACACCTGCTGTGTTCAGACCGCAGACCTGACCTTATGGCAGT 1008
|||||
357 LeuGlyHisProLeuSer 362
|||||
1009 CTGGGTATCTCCATCTCT 1026
|||||
seq_name: gp_ro:BC021634
seq_documentation_block:
LOCUS BC021634 1365 bp mRNA linear ROD 22-JAN-2002
DEFINITION Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA,
complete cds.
ACCESSION BC021634
VERSION BC021634.1 GI:18204912
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 1365)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT Contact: MGC help desk
Email: cgapds@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahney, Erin Helton, Mark Ketteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

FEATURES
source
1..1365
/organism="Mus musculus"
/db_xref="LocusID:27045"
/db_xref="taxon:10090"
/clone="MGC:13825 IMAGE:4008543"
/tissue_type="Mammary tumor metastasized to lung.
MMTV-LTR/Int1 model. Expression driven by an MMTV-LTR
enhancer."
/clone_1lb="NCI-CGAP-Lu30"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
59..1030
/codon_start=1
/product="nitrilase 1"
/protein_id="AAH21634.1"
/db_xref="GI:18204913"
/translation="MLGFTTRPPEHQLCTGYRLRTPVLTCTOPRPTMSSTSMELPL
VAVCOVTPNKOENFKTCALYQEAARAGACVAFLEPAEDFIARNPAPFLLISEPLN
GDLGGVSGLARPCGIMLSLGGHGERGDOMENOKLYNCHVLLNSKSVASVYARKHL
CDEVTGQGPMPRESNTYKPGTLEPVPYKTPAGVGAICITDNKFPPLSLKQAGMEI
LYTSSAFSGVTPAHWEVLLRARATSSQCVVIAAOCGRHETRASVYGSVMVDPWGT
VVARCSGEGGLCTARIDLHFLDOMRHLDPVFOHRRPDLGSLGHPLS"

BASE COUNT 368 a 334 c 331 g 332 t
ORIGIN

alignment_scores:
Quality: 1478.00 Length: 349
Ratio: 4.707 Gaps: 3
Percent Similarity: 89.971 Percent Identity: 80.802

alignment_block:
US-09-357-675C-25 x BC021634 ..
Align seg 1/1 to: BC021634 from: 1 to: 1365

14 AspArgThrLeuArgMetValLeuAlaIleSerSerCysArgThrTyrSe 30
|||
24 GATCGACCCCGGTGATGTGGCTGTGCTTCATCTG..... 64
|||
30 rLeuSerArgArgProArgLeuGlyPheIleThrArgProPheHisArgp 47
|||
65 .....GGCTTCATCACACAGGCTCTCTACCA. 91
|||
47 heLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerValLeu 63
|||
92 .....CTCTGTGTACCGGATACCGATATCTTGAACCCAGTACTT 133
|||

```

```

64 CysAlaGlnProArgProArgAlaMetLeuAlaIleSerSerSerSerCysGly 80
134 TGTACTATAGCCAGCCAGGCCAGAACCATGTGCC...TCAATCACTTCCCTGGGA 180
80 uLeuProLeuValAlaValCysGlnValThiSerThrProAspLysGlnG 97
181 GCTGCCCTGGTGGCTGTGTGCCAGGTAAATCAACCAACCAACCAACGAAG 230
97 LnsAsnPhelysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 113
231 AGAACTTTAAAAACATGTGCTCAGTTGGTTTCAACAGAGCTGCCAGACAGGGT 280
114 AlaCysLeuAlaPheLeuProGluLnaPheAspPheIleAlaArgAsp 130
281 GCTTGGCTGGCTTGTGTGCTCGAGGCAATTTGACTTTATTTGACAGAAACC 330
130 oAlaGluThrLeuHisIleuSerGluProLeuGlyLysLeuLeuGluG 147
331 TGGCGAGACATTACTCGTGTGCCAACCACATGAAATGGGAGATCTTTTGGCC 380
147 LuThrThrGlnLeuAlaArgGluCysGlyLeuThrPheLeuSerLeuGlyGly 163
381 AAATATACCCAGCTTGCCAGGAGAAATGTGAATCTGCTGCTGGGGCGGT 430
164 PheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleThrAsnCy 180
431 TTTCACACAGGCTGGCCAAAGATGGGAGAGAAATCAGAAATCTACAAATTG 480
180 sHisValLeuLeuAsnSerLysGlyAlaValAlaValAlaThrTyArgLysT 197
481 TCATGTCTCTTTGACACAGCAAGGATCAGTAGTGCCCAATTCAGAGAAAG 530
197 hrHisLeuCyAspValGluIleProGlyGlnGlyProMetCysGluSer 213
531 CACATCTGTGCGATGTAGAGATCCAGGTCAGAGGCCGATGAGAGAAGC 580
214 AsnSerThrMetProGlyProSerLeuGluSerProValSerThrProAl 230
581 AACTATATACCAAGCTCGGAGGCATCTTTCAGCCACCTTCACAGCACCCGCC 630
230 agLysLysIleGlyLeuAlaValCysTyAspMetArgPheProGluLeuS 247
631 TGGCAAGGTGGTGTACCAATCTGTTATGACATGGGGTTCCTCGAAACTTT 680
247 exLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyProSerAla 263
681 CTTTGAATTTGGCTCAAGCTGGGGCAAGAAATCTACTATCTTCAGCC 730
264 PheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaAr 280
731 TTTGGATCTCTTACAGGTCCAGCCCACTGGGAGGTGCTCTGCGGGCCCG 780
280 gaIaIleGluThrGlnCysTyArgValAlaAlaAlaGlnCysGlyArgH 297
781 CGGCATTTGAATCTCAGTGTATGTAAATGACACACGCGAGTGGAGGCC 830
297 IsthGlnLysArgAlaSerTyArgLysIleSerMetValValAspProTrp 313
831 ACCATGAAACAAGGCAAGTATATGCCCATAGCATGTGTGTTGACCCGTGC 880
314 GlyThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaAr 330
881 GGCACAGTGTGGCCCGCTGCTCGAGGAGGACAGGCCCTGCTGCTGCTCG 930
330 gIleAspLeuAsnTrpLeuArgGlnLeuAlaArgArgHisLeuProValPheG 347
931 AATTGATCTCACACTTGTACAAAGATCCGCCCAACACCTGCTGTCTTC 980
347 InHisArgArgProAspLeuTyArgLysAsnLeuGlyHisProLeuSer 362
981 AGCAGCCGACACCTGACCTCTATGGCAAGTCTGGTGCATCTCACTCTCT 1027
seq_name: gb_hcg:AC084821

```

```

seq: documentation block:
LOCUS AC084821 211771 bp DNA linear HTG-29-JAN-2002
DEFINITION Mus musculus chromosome 1 clone rp23-395h6, WORKING DRAFT SEQUENCE.
ACCESSION AC084821
VERSION AC084821.24 GI:18390260
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 211771)
TITLE Deschamps,S., Gu,W. and Roe,B.A.
JOURNAL Mus musculus BAC Clone rp23-395h6
REFERENCE
AUTHORS 2 (bases 1 to 211771)
TITLE Deschamps,S., Gu,W. and Roe,B.A.
JOURNAL Direct Submission
Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019 USA
COMMENT
On Jan 29, 2002 this sequence version replaced gi:16076995.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 211771: contig of 211771 bp in length.
*
location/Qualifiers
1. 211771
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-395h6"
/clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT 55622 a 48688 c 49207 g 58231 t 23 others
ORIGIN

alignment_scores:
Quality: 1105.50 Length: 676
Ratio: 3.893 Gaps: 8
Percent Similarity: 42.012 Percent Identity: 37.722

alignment_block:
US-09-357-675C-25 x AC084821 ..

Align seg 1/1 to: AC084821 from: 1 to: 211771

64 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysGsl 80
||||| |||||||:|||||:|||||
975 TGTGTCATTCACAGCGCCAGAACCAATGTC...TCATCAACTCTCGGGA 1021
GCGTGCCTCGGTCGTGTGCCAGGTAATCAACACCAACACACAG 1071
1022 GCGTGCCTCGGTCGTGTGCCAGGTAATCAACACCAACACACAG 1071
80 uleuproleuValAlaValCysGlnValThrSerThrProAspGly 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
97 InAspPheIyThrCysAlaGluLeuValArgGluAlaAarGLeuGly 113
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||
1072 AGAAGCTTTAAACATGTGCTGACTTGCTTCAGAGAGCTGCCAGACG 1121
114 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsp 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```











49539 TGTAAACCCCCCATGTCAGAGCTAATGAAGCACTACAGTCTCTGT 49588  
 233 .....IleGlyLe 235  
 49589 TGAATGGACAGGAGAGATGAGTGGTGTCTTCATTTCAGGTGGCT 49638  
 235 uAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 252  
 49639 AGCAATGCTTATGACATGGCGTCCCTGAACCTTCTTTGAATGGCTC 49688  
 252 lAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 268  
 49689 AAGCTGGGGAGAAATACTTACTTATCTTCAGCCTTGGATCTGTACA 49738  
 269 GlyProAlaHisTrp..... 273  
 49739 GGTCCAGCCCACTGGGAGGTAGAATGCCTTTCATACATATAGGCC 49788  
 273 ..... 273  
 49789 TCCTTTAACCTTACTTTAGTGGGAGCAGAGTGTGACAACCTAAT 49838  
 273 ..... 273  
 49839 CCAAGCACTTCAGTGATGACCAAGCTTCAAGCCAGCATGCTTGACT 49888  
 273 ..... 273  
 49889 AAAAAAGAAAAAATAAGAAAGCTGGGCAAGTGGGGTGCACACCTTAAAT 49938  
 273 ..... 273  
 49939 CCCAGCACTCAGAGGAGGAGAAAGCTGATCTGTGAGTTGAGGCCAGC 49988  
 273 ..... 273  
 49989 CTGGCTACAGAGTGGGTTCAGAGACAGCCAGGGCTACACAGAAACC 50038  
 273 ..... 273  
 50039 TGTCTCAAAAAAACAACCAACCCACACTTCAAGAGTGGGCATNGGT 50088  
 273 ..... 273  
 50089 GGTGCATGCTCTTATCTCAGCACTGTGAGCAAGGAGGAGCAGATCTCT 50138  
 273 ..... 273  
 50139 GTGGGCTTAAGCCAGCAGGGTTACAGAGGAGACCTTATCTCAAAAGTA 50188  
 273 ..... 273  
 50189 CAAACACAAACACTCTCTCTTCTACTGTGATTTCTTCTTTCATCAGTG 50238  
 273 ..... 273  
 50239 GTGAGGAATTCATTCTTACCTAGGTGTGATCATGTAAAGACTTACT 50288  
 273 ..... 273  
 50289 TATTAACTCTAGTGTGTAGTGAAGATTTGGCCACAGTCCCGCTTC 50338  
 273 ..... 273  
 50339 TTGTCACCTTGCCATTAAACCGCTGATGTTCTGAGTCACGACGTCCTC 50388  
 274 .....Glu..ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVal 288  
 50389 ATCTCTCAGGTCTGCTGGGGCCCGCCGCAATTCATTCAGTCTCATGTA 50438  
 289 ValAlaAlaAlaGlnCysGlyArgHisHisGluIleValArgAlaSerTyrG 305  
 50439 ATACGACAGCGCAGTGTGAGCGCCACCATGAACAACAGCAGATTTATGG 50488

305 yHisSerMetValAlaAspProTyrGlyThrValValAlaAlaArgCysSerG 322  
 50489 CCATAGCAGTGGTGGTACCCGTCGGGACAGTGGTGGCCGCTGCTCCG 50538  
 322 lGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGln 338  
 50539 AGGAGCCAGGGCTCTGCTGCTCGAATTCATTCACATTTCTACAAAG 50588  
 339 LeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrG 355  
 50589 ATGGCCCAACACCTGCTGCTTCAGCAGCAGCAGCAGCTGACTTARG 50638  
 355 yAsnLeuGlyHisProLeuSer 362  
 50639 CAGTCTGGTCACTCCTCTCT 50660  
 seq\_name: gb\_htg:AC105589  
 seq\_documentation\_block:  
 LOCUS AC105589 178418 bp DNA linear HTG 09-JAN-2002  
 DEFINITION Rattus norvegicus clone CH230-242B2, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 65 unordered pieces.  
 ACCESSION AC105589  
 VERSION AC105589.1 GI:18092812  
 KEYWORDS HTG: HTGS PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 178418)  
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
 Alsbrooks,S.L., AmaralLunge,H.C., Are,U.R., Banks,T., Barbara,J.,  
 Benton,D., Bimaga,K., Blankenburg,K., Bonin,D., Bouck,J.,  
 Bowls,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
 Burch,P., Burrell,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,  
 Carter,M., Cavazos,S.R., Checko,J., Chavez,D., Chen,R.,  
 Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Geo,U., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
 Hernandez,J., Hernandez,O., Hodgeson,A., Hogues,M., Holloway,C.,  
 Hollins,B., Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,J., Korvah,J.,  
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,  
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabhat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,  
 Ogih,M., Okunou,G., Oragune,N., Oviedo,R., Pace,A., Payton,B.,  
 Peery,J., Perez,L., Peters,L., Pickens,R., Plums,E., Pu,L.L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,  
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshtrli,N.,  
 Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Wiczek,R., Wooden,S.,  
 Watlington,S., Williams,G., Williamson,A., Wiczek,R., Wooden,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,  
 Weinstock,G. and Gibbs,R.  
 TITLE  
 JOURNAL  
 Direct Submission  
 Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

2 (bases 1 to 178418)  
Worley, K.C.  
Direct Submission  
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GNGI

Center clone name: CH230-242B2

----- Summary Statistics

Assembly program: Phrap; version 0.990329first call to

findphraplist

Consensus quality: 132588 bases at least Q40

Consensus quality: 142233 bases at least Q30

Consensus quality: 152072 bases at least Q20

Estimated insert size: 141845; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-1p estimation

Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 65 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 8302: contig of 8302 bp in length  
\* 8303 8402: gap of unknown length  
\* 8403 18472: contig of 10070 bp in length  
\* 18473 18572: gap of unknown length  
\* 18573 24922: contig of 6350 bp in length  
\* 24923 25022: gap of unknown length  
\* 25023 30899: contig of 5877 bp in length  
\* 30900 30999: gap of unknown length  
\* 31000 38320: contig of 7321 bp in length  
\* 38321 38420: gap of unknown length  
\* 38421 43244: contig of 4824 bp in length  
\* 43245 43344: gap of unknown length  
\* 43345 48228: contig of 4884 bp in length  
\* 48229 48328: gap of unknown length  
\* 48329 54595: contig of 6267 bp in length  
\* 54596 54695: gap of unknown length  
\* 54696 58551: contig of 3856 bp in length  
\* 58552 58651: gap of unknown length  
\* 58652 62304: contig of 3553 bp in length  
\* 62305 62304: gap of unknown length  
\* 62305 66587: contig of 4283 bp in length  
\* 66588 66687: gap of unknown length  
\* 66689 70928: contig of 4241 bp in length  
\* 70929 71028: gap of unknown length  
\* 71029 75627: contig of 4599 bp in length  
\* 75628 75727: gap of unknown length  
\* 75728 78372: contig of 2545 bp in length  
\* 78373 78372: gap of unknown length  
\* 78373 80301: contig of 1929 bp in length  
\* 80302 80401: gap of unknown length  
\* 80402 83675: contig of 3274 bp in length  
\* 83676 83775: gap of unknown length  
\* 83776 87173: contig of 3398 bp in length  
\* 87174 87273: gap of unknown length  
\* 87274 90196: contig of 2923 bp in length  
\* 90197 90296: gap of unknown length  
\* 90297 92065: contig of 1769 bp in length  
\* 92066 92165: gap of unknown length  
\* 92166 94612: contig of 2447 bp in length

94613 94712: gap of unknown length  
\* 94713 94712: contig of 403 bp in length  
\* 94714 98755: gap of unknown length  
\* 98756 98855: gap of unknown length  
\* 98856 102543: contig of 3688 bp in length  
\* 102544 102643: gap of unknown length  
\* 102644 105833: contig of 3190 bp in length  
\* 105834 105933: gap of unknown length  
\* 105934 108499: contig of 2566 bp in length  
\* 108500 108599: gap of unknown length  
\* 108600 111672: contig of 3073 bp in length  
\* 111673 111772: gap of unknown length  
\* 111773 114662: contig of 2890 bp in length  
\* 114663 114762: gap of unknown length  
\* 114764 116145: contig of 1383 bp in length  
\* 116146 116245: gap of unknown length  
\* 116246 118147: contig of 1902 bp in length  
\* 118148 118247: gap of unknown length  
\* 118248 120579: contig of 2332 bp in length  
\* 120580 120679: gap of unknown length  
\* 120680 122618: contig of 1939 bp in length  
\* 122619 122718: gap of unknown length  
\* 122719 124539: contig of 1821 bp in length  
\* 124540 124639: gap of unknown length  
\* 124640 126511: contig of 2012 bp in length  
\* 126512 126751: gap of unknown length  
\* 126752 128460: contig of 1709 bp in length  
\* 128461 128560: gap of unknown length  
\* 128561 130592: contig of 2032 bp in length  
\* 130593 130692: gap of unknown length  
\* 130693 132864: contig of 2172 bp in length  
\* 132865 132964: gap of unknown length  
\* 132965 134881: contig of 1917 bp in length  
\* 134882 134981: gap of unknown length  
\* 134982 136682: contig of 1701 bp in length  
\* 136683 136782: gap of unknown length  
\* 136783 139138: contig of 2356 bp in length  
\* 139139 139238: gap of unknown length  
\* 139239 140757: contig of 1519 bp in length  
\* 140758 140857: gap of unknown length  
\* 140858 142558: contig of 1701 bp in length  
\* 142559 142658: gap of unknown length  
\* 142659 144303: contig of 1645 bp in length  
\* 144304 144403: gap of unknown length  
\* 144404 145490: contig of 1087 bp in length  
\* 145491 145590: gap of unknown length  
\* 145591 147171: contig of 1581 bp in length  
\* 147172 147271: gap of unknown length  
\* 147272 148639: contig of 1368 bp in length  
\* 148640 148739: gap of unknown length  
\* 148740 150295: contig of 1556 bp in length  
\* 150296 150395: gap of unknown length  
\* 150396 151505: contig of 1110 bp in length  
\* 151506 151606: gap of unknown length  
\* 151607 152813: contig of 1208 bp in length  
\* 152814 152913: gap of unknown length  
\* 152914 155423: contig of 2510 bp in length  
\* 155424 155523: gap of unknown length  
\* 155524 156880: contig of 1357 bp in length  
\* 156881 156980: gap of unknown length  
\* 156981 158045: contig of 1065 bp in length  
\* 158046 158145: gap of unknown length  
\* 158146 159255: contig of 1110 bp in length  
\* 159256 159355: gap of unknown length  
\* 159356 160417: contig of 1062 bp in length  
\* 160418 160517: gap of unknown length  
\* 160518 161644: contig of 1127 bp in length  
\* 161645 161744: gap of unknown length  
\* 161745 163180: contig of 1436 bp in length  
\* 163181 163280: gap of unknown length  
\* 163281 164483: contig of 1203 bp in length  
\* 164484 164583: gap of unknown length  
\* 164584 166039: contig of 1456 bp in length  
\* 166040 166139: gap of unknown length

```

* 166140 167290: contig of 1151 bp in length
alignment_scores:
    Quality: 1064.00      Length: 836
    Ratio: 3.512          Gaps: 11
    Percent Similarity: 36.244    Percent Identity: 32.057
alignment_block:
US-09-357-675C-25 x AC105589 ..
Align seq 1/1 to: AC105589 from: 1 to: 178418

36 ArgLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuLeu 52
|||||
39092 AGGCTGGGCTTCATCACCAGGCTCTCATCACTCTGCTCCCTTCTGTG 39141
52 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro... 67
|-|-|
39142 TACCGGATACCGGATTTACCTCAATCTCAGTACTTGTACTCAGCCAGGT 39191
67
39192 AACATATTTATGTTCTTGTCTAGACAGTAGTGTCTAAGTCACTGTTTA 39241
67
39242 TTACCCGAATAGTGAGCTGCCATCCACCAACATTTATGCGTGTACTC 39291
67
39292 CAGTAGCTTAGGCTTATTTATGCTCTGAGAGATGAACAAACAGTCGCTA 39341
67
39342 GGGTTGGTGACACACCTCTAATCCAGACTCGGGAACAGCAGAGAG 39391
67
39392 GATCACTAGAGATTAAATCATCCCGCTACATAGAAAGTCTGCTG 39441
67
39442 GGTACAAAAAAGACCGGCTCTCAAGGGGAGGTAAATCAGAAAAAAA 39491
67
39492 AAGTAGTTCCGGCTCCACCCCACTTGAGCTCCCAAGTTTCTCCGG 39541
68
39542 GCACGCGGTAGAAATCCTGAGAAACCTGCTGTCTCATTCACAGGCCAG 39591
70
39592 AGGCATGGCAGAGTCTCATGACTTCTCTGGAGCTGCCCTGGTGGCTG 39641
86
39642 TGTGCCAGGTAACTCAACCAACCAAGCAAGAACTTTAAACGCTGT 39691
103
39692 GCTAGCTGTTCAAGAGGCTACAGAGCTGGCGCTTCCCTGGCTTCT 39741
119
39742 GCGTGAAGCATTTGACTTTATTCACAGAAATCCTGCCAGACATTATACC 39791
136
39792 TGTCCGAACCATGTGATGGGAGCTTTTGGAGCAATATATAGCAGCTTCC 39841
152
152

```

---

```

39842 AGTACAAAGGCTAGGCTGGGGGTGGGGGAGGGGTANATCCCTGTGGG 39891
152
39892 CAGTGTCTCTGAGATTGCCACATAGAAAGGGTAGAGTGTCTCAGTCTCC 39941
153
39942 CCCCTCAGGAATGTGGAATCTGGCTGTCTGGGTGGTTCCACCAACG 39991
167
39992 TGGCCCAAAATGGGAGACAGAAATCCGAAATCTACCATGTCTCGTGGCTTC 40041
184
40042 TGAACAAACCAAGGAGAGACATTCACATCTTCACCTGGCTCTCTAAG 40091
187
40092 GTCACTACACTTCTCTCAGCTAGACTTAACATTTTTCACCTGCTTG 40141
187
40142 GCTTAGGGGTTTTTTTGTGTTTGAATGCTTCTGTCTTACCTAA 40191
187
40192 AAATTACTCTTTGGGAGAAGACTAAGCTGCAGCTCCAGTGTCTACTG 40241
188
40242 AGCATTTTCTCTTACTTATGATCAGTGTGCGCAGTTACAGAGAGAC 40291
198
40292 ATCTATGTGATGTAGAGATCCAGGCTCAGGGCCCTATGAGAAAGCAAC 40341
215
40342 TATACCATGCTCGATGATGCTCTTGAACCACTGTCAAGACACAGCAG 40391
231
40392 CAAGGTAGAGATTGAAAAAATAGAAATATGTGAAAAAGTTGTAAC 40441
235
40442 ATCTCCCTTGTGTAGATGATGAGGATGATTCATGTTACGTTGAC 40491
236
40492 CAGGACAGGAAGATGAGTGGTGTCTTCTCTCTCAGAGGTGCTTACA 40541
237
40542 ATCTGTATGATGATGCGGTCCCTGACCTTCTCTGAATGGGCTCAAGC 40591
253
40592 TGGGCCAATAATCTTACTTATCTTCAAGCTTTTGATCTGTACCGGCC 40641
270
40642 CAGCCACACTGGAGGTAGAGAAATCCTCTTCATTAACATATAGCCCTCT 40691
275
40692 TTAATCTTATTTTAGTAGTGGGAGAGATAGACAAACCTTAACCTTA 40741
275
275

```

```

40742 AGCACTTCAGGGGATAGCCACAGGTTCAAAAGCCAGCATAGTATGATGCTCG 40791
275 .....
40792 AGTAAAAACAGAAAAAATAGAAAGCTGGGCAAGTGGGGGTGCACACCTTT 40841
275 .....
40842 AATCCAGCCCTCAGAGGAGCAAGAGCAAGTGAATGATGATTCAAGACCAG 40891
275 .....
40892 CCTGTCTACAGAGTAGTTCACAGACGACGAGGCTTACACAGAGAAACC 40941
275 .....
40942 GTCTCAAAACAAAACAAAACAAAACAAAAGTAACCCACACACCC 40991
275 .....
40992 AAAATAGTGTGCATGTGGTGTGTGCTTTTATCTCCGACATGTAAGGC 41041
275 .....
41042 AGAGCCCGCAGATCTGTGGGTCCAGGCCAACCAAGTTATACAAAGA 41091
275 .....
41092 GACCTTATATAAAATAATACAAAACAAACACACCCCTCCCTTCTACT 41141
275 .....
41142 TGAAGATTTTCTCCCTCCGATCGTGAGAAATTCGTTCTTAGAT 41191
275 .....
41192 AGGTGTGATCATGTTATAGTACTAATCTAGTATCATATAGTC 41241
275 .....
41242 AATGATGATTTGGGCCACAGTCCGCTTCTGTACCTCGCCTATTAA 41291
276 .....
41292 ATGCTGATGTCTCTAGTACGACGCTCATCTCCTCAGTGTGCTGCGG 41341
279 AAlaArgAlaIleGluThrGlnCysTyrValAlaAlaIleGlnCysG 295
41342 GCCCGGCGCATTTAGTCTCAGTCTAGTATAGCAGACACAGTGTGG 41391
295 YATGHTSHISGluLysArgAlaSerTyrGlnHisSerMetValAlaAsp 312
41392 ACCGACCAATGAGACAGACAGAGTTAGGSCACAGCATGTTGTTTACC 41441
312 GATTCGlyThrValAlaAlaArgCysSerGluLysProGlyLeuCysLeu 328
41442 CTTGGGGACAGTGTGTGGCCAGCTGTTCCAGGGACAGGCTTTGCTT 41491
329 AAlaArgTleAspLeuAsnTyrLeuArgGlnLeuArgHisLeuProVa 345
41492 GCTCGAATTTATCTCCACTTTCTGCAACAGATGGCCACACTTGGCTGT 41541
345 lPhelGlnHisArgArgProAspLeuTyrGlnLysLeuGlnHisProLeus 362
41542 GTTCCAGCATCGCAGACCTTATGTCAGATCTC...CTGCACTCT 41588
362 er 362
41589 CC 41590
seq_name: gp_ov:AF284575
seq_documentation_block:

```

```

LOCUS AF284575 1214 bp mRNA linear VRT 22-JUL-2000
DEFINITION Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.
ACCESSION AF284575
VERSION AF284575.1 GI:9367117
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 1214)
Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
Bleganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.
Crystal structure of the worm NitPhit Rosetta Stone protein reveals
a Nit tetramer binding two Phit dimers
Curr. Biol. 10 (15), 907-917 (2000)
JOURNAL MEDLINE 20414396
REFERENCE
2 (bases 1 to 1214)
Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
Bleganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.
Direct Submission
Submitted (05-JUL-2000) Kimmel Cancer Center, Thomas Jefferson
University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA
FEATURES
Location/Qualifiers
1..1214
/organism="Xenopus laevis"
/db_xref="taxon:8355"
1..1214
/gene="Nit1"
180..1046
/gene="Nit1"
/Note="Nit1: similar to the Nit domains of Caenorhabditis
elegans NitPhit and Drosophila melanogaster NitPhit"
/codon_start=1
/product="Nit protein 1"
/protein_id="AF87104.1"
/db_xref="GI:9367118"
/translation="MAGAHKPLIAVCOMTSTSDKEKNFATCSRLIREAGRRACWFL
PEAFDYIGSIEETLSLAESLHGDTIQRYTOLARCGMLSLGFGHEHKGPMWDORI
NSHVYVNTGTHIVSVYRAHLFDVLDNGVLSRSSSTLPQALIRITSPAGKIG
GVCYDLRPEPSIALAOGACELTTPPSAFTTGTGAAHEVILRRARITGCVYVAAO
TDRAHEKRTSYGHANVYDPMGLVIGCGCEGTICAEIDIFYMERVRDMVWRHRT
DLYKISFNKPD"
BASE COUNT 281 a 305 c 345 g 282 t 1 others
ORIGIN
alignment_scores:
Quality: 968.50 Length: 307
Ratio: 3.905 Gaps: 1
Percent Similarity: 80.782 Percent Identity: 57.980
alignment_block:
US-09-357-675C-25 x AF284575 ..
Align seg 1/1 to: AF284575 from: 1 to: 1214
51 LeuCySProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67
|||||
144 CTGTGTCGCAATATAGGGGTTAGAGCCTGAGTGCATG..... 182
67 oArgProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuV 84
183 .....
84 AlAlaValCysGlnValThSerThProAspLysGlnGlnAsnPhelys 100
:::|||||
205 TTGCGGTGCCAGATCACTTCAACCTTGATAGGAGAAATTTTCGCC 254
101 ThnCysAlaGluLeuValArgGluAlaAlaArgLeuGlyValCysLeuAl 117
|||||
255 ACGGTTCGCGGCTGATCGGAGGCTGGGGCGCTGCGCTTGATGAT 304

```



```

176 LysIleTyraNcysHisValIleuLeuAsnSerLysGlyAlaValAl 192
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84896 AAAATCTTCAACGCTCATGTTTCTCAAGAGAAAGGCAACAGACAGC 84945
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
192 aATTtYArGLyLThrHisLeuCyAspValGluIleProGlyGlnGlyP 209
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84946 AGATATACAGAAAGCTGCACATGTTGATGTTACACTTAA...GAGGTT 84992
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
209 roMetCyGusSerAsnSerThrMetProGlyProSerLeuGlnSerPro 225
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84993 GCCTACGCGCATCAGATACAGTTACGCGCGGATACCTCTTGACGCCCA 85042
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
226 ValSerThrProAlaGlyLysIleGlyLeuAlaValCysTyraSpmetAr 242
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85043 GTGAGCAGCTCAGATTGGCCAGATAGGCTTCAGATTGTCACAGACCTGC 85092
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
242 gPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuT 259
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85093 TTTTGGTGCACCGCGCTGCTGCACAGAAAGCTGGCCCAATTGTTTAA 85142
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
259 hTrYrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 275
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85143 CATACCATCCGCAATTCATACATACGCAACGCGTAAAGCGCACTGGGAATC 85192
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyraValAlaAlaAl 292
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85193 CTCCTGGGGCCAGAGCCATAGACATACATGCTTGTGCTGGCTGGCGC 85242
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
292 aGlnCysGlyArgHisHisIleGluLysArgAlaSerTyrglyHisSerMetV 309
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85243 TCGATGTGGTGGCGCAACCAAGACAGAGATTGGCGCCACAGCATGA 85292
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
309 aIValAspProTrpGlyThrValValAlaArgCysSerGluGlyPro... 324
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85293 TCGTATCCCTCGGGGAACGTTACTGCTACTGACAGCGAGAGAGCTT 85342
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
325 GlyLeuCysLeuAlaArgIleAspLeuAsnTyrlLeuArgIleuArgAr 341
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85343 GAAATATAGCAGCGCGGAGTGAGACTTCCGCTTCATCTCGTATCA 85392
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 gHisLeuProValPheGlnHisArgArgProAspLeuTyr 354
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85393 GAACTATCCCTCGCTGCACATCTGTCGAAACGACATCTAC 85432
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
seq_name: gbl_in:AC093121
seq_documentation_block:
LOCUS AC093121 168250 bp DNA linear INV 11-AUG-2001
DEFINITION Drosophila melanogaster 3L BAC RP98-11J11 (Roswell Park Cancer
Institute Drosophila BAC library) complete sequence.
ACCESSION AC093121
VERSION AC093121.1 GI:15148095
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 168250)
Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Taber,P., Williamson,A., Homs,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Aylee,M.A., Scott,G.S., Worley,K.W., Amaratilake,P.G., Brando,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Ferriera,S., Gary,N.D.S., Houck,J.,
Hoslin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,M., Mettel,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,Y.,
Schaefer,R., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL

COMMENT

Wheeler,D., Weinstein,G., Gibbs,R. and Venter,J.C.  
Direct Submission  
Unpublished  
2 (bases 1 to 168250)  
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Isman,F.R., Allen,C.,  
Albrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Blincoe,K., Blankenburg,K., Bonnin,D., Bouch,J.,  
Bowling,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dethorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Hollway,C.,  
Hollins,B., Homs,F., Howard,S., Huber,D., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,X., Johnson,R., Jolivet,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kralovic,D., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
LouiSeged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,  
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,M., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
Ogun,M., Okunolu,G., Oregunye,N., Olickson,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojiboken,I., Rolle,M.,  
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoostatli,N.,  
Stinson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalob,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Submitted (11-AUG-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2

clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT.

FEATURES Location/Qualifiers

1..168250  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="3L"  
/clone="RP98-11j11"

BASE COUNT 47670 a 36872 c 37397 g 46311 t  
ORIGIN

## alignment\_scores:

Quality: 715.50 Length: 364  
Ratio: 2.944 Gaps: 6  
Percent Similarity: 66.758 Percent Identity: 42.857

## alignment\_block:

US-09-357-675c-25 x AC093121/rev ..

Align seg 1/1 to reverse of: AC093121 from: 1 to: 168250

```

17 LeuArgMetValLeuAlaIleSerSerCysArgThrTySerLeuSerArg 33
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105593 CTTTCCCTGGTGTGTGTACACACTGTTTGCACATCTCTCGGAAACG 105544
   ||| |
33 gArgPro..... 35
105543 CCGTAACCTAGGGGATGCTAAAGACGACGCACTAAACACTAGGTA 105494
   36 .....ArgLeuGlyPheIleThrArg 42
105493 AAAATATTGATATCATCTTACGACGTAACGATTCGAAATTTTCCACAG 105444
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
43 ProProHisArgPhe.LeuSerLeuLeuCysProGlyLeuArgIleProG 59
   ||| |
105443 CCGGCGACCCGTTCAATGCTCACTAGTTAATACACGAGA.....C 105400
   ||| |
59 InLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSer 75
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105399 GCACGATTGTGATGCCATTCATCAGCAGTACGACGATGTCGCTCA 105350
   76 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSer 92
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105349 AAGAGCAAGCATCAAGTCCGACCATTCCTGCTGATGATCCGCTTAC 105300
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
92 rProAspLysGlnGlnAsnPhelysthCysAlaGlnLeuValArgGln 109
   | ||||| |
105299 CAGCGACAAGCGCGCTAATCTTACCAAGTATGATAGCTAGTATGAGG 105250
   | ||||| |
109 IaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPhe 125
   | ||||| |
105249 CCAAGTCACAAACGCTGCTCTTCTGCTGAGTGTGCTGACTTT 105200
   | ||||| |
126 IleAlaArgAspProAlaGlnThrLeuHisLeuSerGluProLeuGly 142
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105199 GTGGCGCAGACCGCGACCAACTATTCAGCTCTCCGAGGCGCTTGACGG 105150
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
142 YLysLeuLeuGlnGluTyThrGlnLeuAlaArgGluCysGlyLeuTyr 159
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105149 CGAGTTAATGGCGAGTACCGGAAATGGCAAGTGCMAAATAATTTTGA 105100
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
159 euserLeuGlyGlyPheHisGlnArgGlnAspTyrPheGlnThrGln 175

```

```

105099 TTTCCCTGGTGGCGTGCACGACGCAACGAT.....CAA 105065
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
176 LysIleTyrAsnCysHisValLeuAsnSerIysGlyAlaValAla 192
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105064 AAAATCTTCAACGCTCATGTTTGTCTCAACGAAAGGGAACCTAGCAGC 105015
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
192 aThrTyArgIsthrHisLeuCysAspValGlnIleProGlnGlnGly 209
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105014 AGTATACGAAAGCTGCACATGTTGATGTTACGACTAAA...GAGGTTTC 104968
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
209 rometCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 225
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104967 GCCTACGCGAATTCATACATACATACGCGCGGATCTGCTTGCAGCGCCA 104918
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
226 ValSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetAr 242
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104917 GTGAGCACTCCAGTTGGCCAGATAGCGCTTCAGATTCTCTACGACCTGCG 104868
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
242 gPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyValGluIleLeu 259
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104867 TTTTGTGAGCGCGGCGGTCTCTCAGAGAGCTGGGTCCCAATTTGTTAA 104818
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
259 hTyrProSerAlaPheGlySerIleThrGlyProAlaHisTyrPheVal 275
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104817 CATACCATCCGCAATTCATACATACGACCGGTAAGCCACTGGGAAATTC 104768
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValAlaAlaAla 292
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104767 CTCCTGGCGGCGCAGAGCCATAGAGACTCATGCTTGTGTGCGCGCGGC 104718
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
292 gLysGlnGlyArgHisIleGluLysArgAlaSerTyrGlyHisSerMetV 309
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104717 TCAGTTAGTTGGCACAACAGAACGACAGAGTTGGGGGCCACACATGGA 104668
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
309 aValAspProTyrPheGlyThrValValAlaArgCysSerGluCylPro... 324
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104667 TCGTTAGCCCTCGGGGAACGTAAGCTGCTGACTGACGACGACGACGACTT 104618
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
325 GlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArg 341
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104617 GATATAGCGACGCGCGAGGTGACCTTCCGTCCTCAATCTCTGTATCA 104568
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 gHisLeuProValPheGlnHisArgArgProAspLeuTyr 354
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104567 GACCATGCCCTGCTTGCAGACATCGTCGAAACGACATCTTAC 104528
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```





OM of: US-09-357-675c-25 to: N.Geneseq\_032802:\* out\_format : pfs  
Date: Apr 29, 2002 10:22 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODE=frame+ p2n.model -DEF=yih  
-Q/cgn2.1/USFTO.spool/US09357675/runatc.29042002\_091757\_3254/app\_query.fasta.1.1208  
-Db=N.Geneseq\_032802 -QPM=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MAPFIX=blosome2  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09357675.@CGN1\_1.412 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPHY -WAIT -THREADS=1

## Search information block:

Query: US-09-357-675c-25  
Query length: 362  
Database: N.Geneseq\_032802:\*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 338.150000

## score list:

Sequence	Strd Orig	ZScore	EScore	len	! Documentation	..			
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +		1919.00	2947.61	7.8e-15					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +		1806.00	2773.24	4.0e-14					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +		1067.00	1632.70	1.4e-82					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ30398 +		968.50	1480.40	4.1e-74					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAZ25458 +		715.50	1078.53	1.0e-5					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI2224 +		715.50	1073.71	1.9e-5					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI2225 +		710.50	1079.60	8.7e-5					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ46101 +		495.00	750.69	1.8e-32					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAZ25456 +		485.50	732.89	1.8e-32					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ25457 +		481.00	726.45	4.1e-32					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +		404.00	566.81	3.2e-23					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ21607 +		404.00	550.40	2.6e-22					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ31490 +		404.00	536.01	1.7e-21					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ71176 +		392.00	585.98	2.7e-24					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI3787 +		387.00	584.90	3.1e-2					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI3786 +		386.50	572.01	1.6e-2					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI5102 +		385.00	567.75	2.8e-2					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +		301.00	456.24	4.6e-17					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ68533 -		292.50	378.11	1.0e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ68533 -		291.00	436.78	5.6e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ7312 +		291.00	436.78	5.6e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ7312 +		283.50	426.17	2.2e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ82550 -		283.50	372.36	2.2e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ82550 -		276.50	406.11	2.8e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAZ82413 +		276.50	405.44	3.1e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAZ82413 +		270.00	401.45	5.1e-14					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ50003 +		270.00	401.45	5.1e-14					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ50003 +		262.00	369.15	4.0e-12					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAZ52205 +		262.00	369.15	4.0e-12					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAZ52205 +		248.00	351.51	3.1e-11					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:AAZ30056 -		229.00	346.86	5.7e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAZ82413 +		229.00	346.86	5.7e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAZ82413 +		220.50	308.45	7.8e-09					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAZ7312 -		220.50	308.45	7.8e-09					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ56993 +		218.00	325.00	9.4e-1					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAZ56993 +		218.00	309.58	6.8e-0					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ47871 +		217.50	319.80	1.8e-09					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56975 +		214.50	311.35	5.4e-09					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56975 +		213.50	309.80	6.6e-09					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56992 +		213.50	309.80	6.6e-09					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ40019 +		212.50	312.99	4.4e-09					
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56971 +		212.50	308.26	8.0e-09					

/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56972 + 212.50 308.26 8.0e  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56993 + 212.50 308.26 8.0e  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56994 + 212.50 308.26 8.0e  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56973 + 211.50 308.71 9.8e  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AAZ56968 + 211.50 306.71 9.8e

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101

## seq\_documentation\_block:

ID AAZ46101 standard; DNA; 1416 BP.  
XX  
AC AAZ46101;  
XX  
DT 05-MAY-2000 (first entry)  
XX  
DE Nucleotide sequence of the coding region of NIT1 gene.  
XX  
KW NIT1 gene; nitrlase; tumour suppressor gene; FHIT; chromosome 3p14.2;  
KM FRA3B; cancer; genome allele inactivation; ss.  
XX  
OS Homo sapiens.  
OS Mus sp.  
OS Drosophila melanogaster.  
OS Caenorhabditis elegans.  
XX  
FH Key Location/Qualifiers  
FT CDS 3..1415  
FT /\*tag= a  
FT /product= (pos: 18..20, aa: Xaa)  
FT /product= (pos: 1179..1181, aa: Gln)  
FT /product= (pos: 1182..1184, aa: Ala)  
FT /note= "contains 7 internal stop codons; Xaa is an unspecified amino acid"

WO2000033685-A2.

27-JAN-2000.

20-JUL-1999; 99WO-US16366.

20-JUL-1998; 98US-0093350.

(UYE-) UNIT JEFFERSON THOMAS.

Croce CM;

WPI; 2000-171195/15.

P-PSDB; AAY68739.

Novel nitrlase homologs used as diagnostic and therapeutic reagents for the detection and treatment of cancer -

Claim 6; Fig 6; 25pp; English.

The present sequence represents the coding region of human, murine, Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The human and mouse NIT1 genes are members of an uncharacterised mammalian gene family with homology to bacterial and plant nitrlases. The tumour suppressor gene FHIT in D. melanogaster and C. elegans code for fusion proteins in which the Phit domain is fused with a Nit domain. In mouse and humans, FHIT and NIT are encoded by two different genes, localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The human FHIT gene at chromosome 3p14.2, spanning the constitutive chromosomal fragile site FRA3B, is often altered in most common forms of human cancer. The NIT1 protein overcomes the mutated inactivation of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives and analogues of them, and antibodies are used as diagnostic and therapeutic reagents for the detection and treatment of cancers.

Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;

alignment\_scores:

Quality: 1919.00 Length: 362  
Ratio: 5.316 Gaps: 0  
Percent Similarity: 99.724 Percent Identity: 99.724

alignment block:

US-09-357-675c-25 x AA246101 ..

Align seg 1/1 to: AA246101 from: 1 to: 1416

```

1  ProLeuAlaAlaAla**LeuAlaProAspArgProAspArgThrLeu 17
   |||
3  CCACCTGCGCGCCCTNTGCTGCTCCAGACCGCCCTCGGATCGGACCT 52
   |||
17 uArgMetValLeuAlaIleSerSerCysArgThrTyrrSerLeuSerArg 34
   |||
53 GCGAATGGTTTGGCTATATCTTTCATGTAGAGACTTACTCCATCCGCG 102
   |||
34 rGProArgLeuGlyPheIleThrArgProProHisArgPheLeuSerLeu 50
   |||
103 GCGCGCGGCTGGGCTTCATCACCAGGCCCTCCACAGATTCCTGTCCTT 152
   |||
51 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67
   |||
153 CTGTGTCCTGGACTCGGATACCTCACTCACTCACTTGTGCTAGCC 202
   |||
67 oArgProArgAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuY 84
   |||
203 CAGGGCCAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCT 252
   |||
84 aAlaValAlcysGlnValThrSerThrProAspLysGlnIleAsnPhelys 100
   |||
253 TGGCTGTGTGCCAGGTAACTGACGCGCAGACAGACAGACAGAACTTAA 302
   |||
101 ThrCysAlaGlnLeuValArgGlnAlaAlaArgLeuGlyAlaCysLeuAl 117
   |||
303 ACATGTGCTGAGCTGTGCTGAGAGGCTGCGACAGCTGGGCTGCTGCG 352
   |||
117 aPheLeuProGlnAlaIlePheAspPheIleAlaArgAspProAlaGluThrL 134
   |||
353 TTTCTGCTGCTGAGGCAATTGACTTGCATTCGACGAGGACCTGCGAGAGCC 402
   |||
134 euHisLeuSerGlnProLeuGlyGlyLysLeuLeuGlnGluThrThrGln 150
   |||
403 TACACCTCTGTAAACCACTGGGTGGGAACTTTGGAAATAACACCCAG 452
   |||
151 LeuAlaArgGlnCysGlyLeuThrPleuSerLeuGlyGlyPheHisGlnLur 167
   |||
453 CTGGCCAGGGAATGTGACTCTGGCTGTCTGCTGCTGCTGCTGCTGCTG 502
   |||
167 gGlyGlnAspTrpGlnGlnThrGlnLysIleTyrrAsnGlyHisValLeuL 184
   |||
503 TGGCCAAAGCTGGAGAGACTCAGAAATACTAACATTTGTCACGTGCTGC 552
   |||
184 euAsnSerLysGlyAlaValAlaIleThrTyrrArgLysThrHisLeuGly 200
   |||
553 TGAACAGCAAAAGGGGAGTGTAGTGGCCACTTACAGAGAACACATCTGTCT 602
   |||
201 AspValGlnIleProGlyGlnGlyPrometCysGlnSerAsnSerThrMe 217
   |||
603 GACGTAGAGATTCAGGGGAGGGGCTATGTGTGAAGCACTTACCTACAT 652
   |||
217 tProGlyProSerLeuGlnSerProValSerThrProAlaGlyLysIleG 234
   |||
653 GCGTGGGCCCAAGTCTTGAAGTCACTGTGACGACACACAGAGGCAAGATTG 702
   |||
234 lYleuAlaValAlcysTyrrAspMetArgPheProGlnLeuSerLeuAlaLeu 250
   |||
703 GTCTAGCTGTCTGTATGACATGCGGTTCCCTGAACTCTCTCTGCGCATTTG 752
   |||
251 AlaGlnAlaGlyAlaGlnIleLeuThrTyrrProSerAlaPheGlySerL 267
   |||
753 GCTCAAGCTGAGCAGAGACTACTTACTTACTTACTTACTTACTTACTTACT 802

```

```

267 eThrGlyProAlaHisTrpGlnValLeuLeuAlaArgAlaIleGluTr 284
   |||
803 TACAGGCCAGCCCACTGGAGAGTGTGCTGGCGGGCCCTGCTATCGA 852
   |||
284 hGlnCysTyrrValValAlaAlaIleGlnCysGlyArgHisHisGlyLys 300
   |||
853 CCAGTGTATGTAGTGGCAGCAGCACTGTGTGAGAGCCACCATGAGA 902
   |||
301 ArgAlaSerTyrrGlyHisSerMetValValAspProTrpGlyThrValVa 317
   |||
903 AGAGCAAGTTATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGT 952
   |||
317 lAlaArgCysSerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeuA 334
   |||
953 GCGCCGCTCTCTGAGAGGGCCAGGCTCTGCTTGGCCCAATAGACCTCA 1002
   |||
334 snTyrrLeuArgGlnLeuAlaArgArgHisLeuProValPheGlnHisArgG 350
   |||
1003 ACTATCTGGCAGAGTGGCCGCGACACCTGCTGTGTCCAGACCGCAG 1052
   |||
351 ProAspLeuTyrrGlyAsnLeuGlyHisProLeuSer 362
   |||
1053 CCTGACCTTATGGCAATCTGGGTGTCACCCACTGTCT 1088

```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAFI6257

seq\_documentation block:

ID AAFI6257 standard; cDNA: 1382 BP.

XX AAFI6257;

DT 13-MAR-2001 (first entry)

DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.

XX Human: prostate cancer: prostate cancer antigen: detection; diagnosis;  
 KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;  
 KW vulnary; gastrointestinal; nephrotoxic; antilinfective; gynaecological;  
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
 KW wound; infectious disease; ss.

XX Homo sapiens.

XX MO20005174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;

XX WPI: 2000-587513/55.

XX P-PSDB: AAB57054.

XX Prostate cancer associated gene sequences, referred to as prostate

XX cancer antigens, useful for treatment, prevention, and diagnosis of

XX disorders such as prostate cancer -

XX Claim 1; Page 1124; 2338pp: English.

XX AAFI5566 to AAFI6505 encode the human prostate cancer associated

XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.

XX The prostate cancer antigens can have neuroprotective, cytosolic,  
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
 CC nephrotoxic, antilinfective, gynaecological and antibacterial activities,  
 CC and can be used in gene therapy. The prostate cancer antigen  
 CC polynucleotides may be used for detection of prostate cancer, chromosome



PR 05-AUG-1997; 97US-0055386.  
 PR 05-AUG-1997; 97US-0055311.  
 PR 18-AUG-1997; 97US-0055970.  
 PR 18-AUG-1997; 97US-0055986.  
 PR 19-AUG-1997; 97US-0056385.  
 PR 19-AUG-1997; 97US-0056366.  
 PR 19-AUG-1997; 97US-0056557.  
 PR 19-AUG-1997; 97US-0056370.  
 PR 19-AUG-1997; 97US-0056371.  
 PR 19-AUG-1997; 97US-0056563.  
 PR 19-AUG-1997; 97US-0056731.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Brewer LA, Ebner R, Ferrle AM, Greene JM, Janat F, Ni J;  
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;  
 XX  
 DR WPI: 1999-167452/14.  
 DR P-PSDB: AAY10877.  
 XX  
 PT New isolated human genes encoding secreted polypeptides - useful for  
 PT diagnosis and treatment of pathological diseases  
 XX  
 PS Claim 3; Page 265-266; 331pp; English.  
 XX

CC The specification describes secreted proteins and their corresponding  
 CC polynucleotides which are useful for preventing, treating or ameliorating  
 CC medical conditions, e.g. by protein or gene therapy. Pathological  
 CC conditions can also be diagnosed by determining the amount of the  
 CC secreted polypeptides in a sample or by determining the presence of  
 CC mutations in the polynucleotides. Specific uses are described for each  
 CC of the products, based on which tissues they are most highly  
 CC expressed in, and include developing products for the diagnosis or  
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental  
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,  
 CC diseases of the immune system, autoimmune diseases, hepatic and renal  
 CC diseases, diabetes, inflammation, allergies, ischemic shock, Alzheimer's  
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,  
 CC prostate diseases, asthma, disorders involving osteoclasts such as  
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or  
 CC thymus, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC  
 XX  
 SQ Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;

alignment\_scores:  
 Quality: 1067.00 Length: 207  
 Ratio: 5.205 Gaps: 0  
 Percent Similarity: 99.034 Percent Identity: 99.034

alignment\_block:  
 US-09-357-675C-25 x AAX30398 ..

Align seg 1/1 to: AAX30398 from: 1 to: 1203

156 GYLeuTRPLeuSerLeuGlyPheHisGluArgGlyGlnAspTrpG 172  
 ||||||||||||||||||||||||||||||||||||||||||||  
 3 GGACTCTGGCTCTCCTGGGTGGTTCATGACGCGGCAAGACGCGGA 52  
 172 uGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyA 189  
 ||||||||||||||||||||||||||||||||||||||||||||  
 53 CGAGACTCAGAAATCTACAAATTGTCACGCTGCTGGAACAG.AAAGGGG 101  
 185 JalaValAlaThrTyrArgLysThrHisLeuCysAspValGluLeuPro 205  
 ||||||||||||||||||||||||||||||||||||||||||||  
 102 CAGTGTGGCCACTTACAGAGAACACATCTGTGACCTAGAGATTCCA 151  
 206 GYglnGlyPheMetCysGlnSerAsnSerThrMetProGlyProSerLe 222  
 ||||||||||||||||||||||||||||||||||||||||||||  
 152 GGGCAGGGGCT.ATGTGTAAAGCAACTTACCAATGCTGGGCGCCAGTCT 200  
 222 uGlnSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCysT 239

||||||||||||||||||||||||||||||||||||||||||||  
 201 TGAGTCACCTGTACAGACACAGCAGGCAAGATTGGTCTACCTGTCTGCT 250  
 239 YrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnIaGlyAla 255  
 251 ATGACATGTGGGTTCCTTGAATCTCTCTGGCAATGGCTCAAGCTGGAGCA 300  
 256 GlnIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaH 272  
 301 GAGATACCTACTATCCTATCCTACGCTTTTGGATCATTTACAGGCCAGCCCA 350  
 272 sTRPGLuValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValY 289  
 ||||||||||||||||||||||||||||||||||||||||||||  
 351 CTGGGAGGTGTGCTGCGGGCCGCTGTATCGAAACCCAGTGTATGTAG 400  
 289 AlAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGly 305  
 ||||||||||||||||||||||||||||||||||||||||||||  
 401 TGGCAGCAGCACAGTGTGGACCCACCATGAGAAAGAGACCAAGTTATGGC 450  
 306 HisSerMetValValAspProTrpGlyThrValAlaAlaArgCysSerG 322  
 ||||||||||||||||||||||||||||||||||||||||||||  
 451 CACAGATGTGGTGTAGACCCCTGGGGAAACAGTGTGGCCGCTGCTCTGA 500  
 322 uGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnL 339  
 ||||||||||||||||||||||||||||||||||||||||||||  
 501 GGGGCGAGGCTCTGCTTGGCCGAAATAGACCTCAACTATCTGCGACACT 550  
 339 euArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 355  
 ||||||||||||||||||||||||||||||||||||||||||||  
 551 TCGCGCGACACTGCTGTGTTCAGCACCGCAGGCTGACCTTATGGC 600  
 356 AsnLeuGlyHisProLeuSer 362  
 ||||||||||||||||||  
 601 AATCTGGTCACCCACTGTCT 621

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT: AAD25458

seq\_documentation\_block:  
 ID AAD25458 standard; cDNA: 1214 BP.

XX AAD25458;

XX 26-MAR-2002 (first entry)

XX Xenopus laevis Nrl1 cDNA.

XX Frog; Nrl1; cytosolic; neuroprotective; cellular pathway; therapy;

XX apoptosis; proliferative disorder; degenerative disease; ss.

XX Xenopus laevis.

XX Key Location/Qualifiers

XX CDS 180..1046

XX /product= "Frog Nrl1 protein"

XX MO200187958-A2.

XX 22-NOV-2001.

XX 15-MAY-2001; 2001WO-US15664.

XX 16-MAY-2000; 2000US-204713P.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Croce C, Brenner C, Pekarski Y;

XX WPI: 2002-082984/11.

XX P-PSDB: AAE15792.

XX Isolated cDNA encoding human, mouse, frog and yeast Nrl2 proteins,

XX useful to find molecules that mimic or antagonize Fhl1 interaction for

PT the treatment of proliferative or degenerative diseases  
XX  
PS Claim 21; Page 60; 61pp; English.

The invention relates to isolated nucleic acids comprising a fully defined cDNA nucleotide sequence encoding human, *Xenopus laevis* and mouse Nlt2 proteins. Nlt and Flt1 proteins are encoded as fusion proteins in invertebrates and as separate polypeptides in vertebrates. Nlt and Flt1 interact physically and functions in same cellular pathways. Molecules which bind Nlt2 and mimic or antagonise Flt1 interaction are used to treat diseases in which activity of Nlt2 protein is altered in a mammal. Flt1 mimics induce apoptosis and are particularly useful to treat proliferative disorders, whilst Flt1 antagonists promote cell proliferation and are particularly useful to treat degenerative disease. The present sequence is frog Nlt1 cDNA.

```
alignment_scores:
  Quality: 968.50      Length: 3077
  Ratio: 3.905         Gaps: 1
Percent Similarity: 80.782  Percent Identity: 57.980
```

```
alignment_block:
US-09-357-675C-25 x AAD25458 ..
```

Align seg 1/1 to: AAD25458 from: 1 to: 1214

```

51 LeuCYsProGlyLeuAnGIIeProGluIneValLeuCYsAlaGlnPr 67
144 CTTGTTCCGGATATAGGGGTTAGAAAGCCTGATGCATAT..... 1822
67 oARgProARgAlaMeAlaIleSerSerSerCysGluLeuProLeu 84
183 .....GTTGGTCCCAAGGCCCTGA 204
84 aAlaValCYsGlnValThrSerThrProAspGlyGlnGlnAsnPhelYs 100
:::|||||:::|||||:::|||||:::|||||:::|||||
205 TTCCCGGTCGCAAGATGACTTCAACCTCTGATTAAGAGAAAGATTTCCGC 254
101 ThrCYsAlaGluLeuValArgGlnAlaAlaIleArgLeuGlyAlaCYsLeuAl 117
|||||:::|||||:::|||||:::|||||:::|||||
255 ACCTGTTTCGGGGCTGATCCGGAGGCTCGGGGGCTCGCGCTTGCAATGGr 304
117 aPhelLeuProGluAlaIleAspPheIleAlaArgAspProAlaGluThrl 134
305 GTTTCGCGCGAAGCCTTGACTATATGCGGGGCAACATTGAGAGACGC 354
134 euhIstLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGln 150
|||:::|||||:::|||||:::|||||:::|||||
355 TGAAGTCGGTGAGTCTCTCATGAGGGGACACCATTCAGGGTTACACCCA 404
151 LeuAlaArgGlyCYsGlyLeuTyrLeuSerLeuGlyGlyPheHisGluAr 167
|||||:::|||||:::|||||:::|||||:::|||||
405 CTGCGCCAGGAGTGGGCTCTGGCTTCCCTCGGGGGGATTCATATAGAA 454
167 gGlyGlnAspTyrPcIuGlnThrGlnLysIleTyrAsnCYsHisValLeu 184
:::|||||:::|||||:::|||||:::|||||:::|||||
455 AGGACCCAACTGGGACACGAGCAACGCATTTCCATTTCTCAGCTGGTTG 504
184 euhSnSerLysGlyAlaValAlaIleAsnTyrArgLysThrHisLeuCYs 200
:::|||||:::|||||:::|||||:::|||||:::|||||
505 TGGACAAACACAGGGCACATATGATGCTGATCGGCAAGGCTCACCTGTTT 554
201 AspValGluIleProGlyGlnGlyProMetCYsGluSerAsnSerThrMe 217
|||||:::|||||:::|||||:::|||||:::|||||
555 GAGGTAGACTTCGAGANTGAGTGCACATCAAGAGAGCAGTTCCACCT 604
217 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 234
:::|||||:::|||||:::|||||:::|||||:::|||||
605 CCCCAGGACAGAGCTTATTCGCCCATCATCTTCCACGACGAGAAATTTG 654

```

```

234  lYleuAlaValcYstrYrYsrpMeLarPheProGluLeuSerLeuAlaLeu 250
      |||.....|.....|.....|.....|.....|.....|.....|.....|
655  GCCCTGGGGGTGGTTACAGCCCTCCGCTTCCAGAAATTCCTCTGGCTCG 704
      |||.....|.....|.....|.....|.....|.....|.....|.....|
251  AlAGlnAlaGlyAlaGluLeuLeuThrTyProSerAlaPheGlySerIl 267
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
705  GCCCAAGACAGAGGACGAACTTCTCACTTACCCTTCTGGCTTACCCTCAC 754
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
267  eThrGlyProAlaHisTrpGluValLeuLeuArgAlaAlaAlaIleGluT 284
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
755  TACTGGTCGGCACCATTGGAGCGTTCCTCGAGACCCGTGCCAATGAAA 804
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
284  hGlcCysTrpValValAlaAlaAlaIaGlnCysGlyArgHisIleGluLys 300
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
805  CCCAGTGCCTACGTAATTGGCAGCGGCACAGACAGACAGACACATATGAAAG 854
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
301  ArgAlaSerTyTrGlyHisSerMetValValAspProTrpGlyThrValVa 317
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
855  AGGAGCGCTCTTAATGGTCACGCGCTATGGTGTAGACCCGTGGGGCTCGTAT 904
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
317  lAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAla 334
      .....|.....|.....|.....|.....|.....|.....|.....|
905  TGGCCCAATGCCACAGGAAGAACAGAAATATGTTATCTGTGAGATTGCAATTC 954
      .....|.....|.....|.....|.....|.....|.....|.....|
334  snTyTrLeuArgGluLeuLeuArgArgHisLeuProValAlaPheGlnHisArgArg 350
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
955  CCTACTATGGAGCGGTGAGAGCGGGGAGCAATGCCGGTGTGAGAGCGACCGCAGG 1004
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
351  ProAspLeuTyTrGlyAsnLeu 357
      |||||.....|.....|.....|.....|.....|.....|.....|.....|
1005  ACTGATCTGTATGGGAAAAATC 1025

```

seq\_name: /SIDSL/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL12224

```

seq_documentation_block:
ID  ABL12224 standard; cDNA; 3548 BP.
XX
XX  ABL12224;
AC
XX
XX  26-MAR-2002 (first entry)
DT
XX
DE  Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.
XX
XX  Drosophila; developmental biology; cell signalling; insecticide;
KW  pharmaceutical; gene; ss.
XX
XX  Drosophila melanogaster.
OS
XX
XX  WO200171042-A2.
PN
XX
XX  27-SEP-2001.
PD
XX
XX  23-MAR-2001; 2001WO-US09231.
PF
XX
XX  23-MAR-2000; 2000US-191637P.
PR
XX  11-JUL-2000; 2000US-0614150.
PP
XX
XX  (PEKE ) PE CORP NY.
PA
XX
XX  Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX  WPI; 2001-656860/75.
DR
XX  P-PSDB; ABB68121.
DT
XX
XX  New isolated nucleic acid detection reagent for detecting 1000 or more
PT  genes from Drosophila and for elucidating cell signalling and cell-cell
PI  interactions -
XX
XX
PS  Claim 1; SEQ ID NO 31154; 21pp + sequence listing; English.
XX
XX
CC  The invention relates to an isolated nucleic acid detection reagent
CC  capable of detecting 1000 or more genes from Drosophila. The invention is
CC  useful in developmental biology and in elucidating cell signalling and

```

CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from  
CC at [ftp://www.ncbi.nlm.nih.gov/pubmed/123456789](http://www.ncbi.nlm.nih.gov/pubmed/123456789).

XX Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;

SQ Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;

[illegible]

**alignment\_block:**

US-09-357-675C-25 x ABL12224 . .

Align seg 1/1 to: ABL12224 from: 1 to: 3548

```

17 LeuArgMetLeuAlaIleSerSerCysAlaGhrTrpIleSerLeuSerAr 33
||| :::::::::::::::::::: ||| ::|||
931 CTTTCCTCGGTGTGTGTGTACACACTTGTGGCACTTCTGTGGGAACG 980
33 gatPro..... 35
|||
981 CCCTAAGCTAGGGGAATGTAAAGAGACGCGCACTTAACCAACTAGTA 1030
36 .....ArgLeuGlyPheIleThrArg 42
|||::: ::::
1031 AAAATGAATTCATACATCTTACGACGTAACGATTCGAAATTTTCCACAG 1080
43 ProProHisArgPhe..LeuSerLeuLeuCysProGlyLeuArgIleProG 59
||| ||||| ::||| |||
1081 CCGGCGACCCGTTTCAATGTCAACTGTAGTTAATACACAGNA.....C 1124
59 IndSerValLeuCysAlaGlnProArgProAlaIleMetAlaIleSer 75
|||::: :::: ||| |||:::
1125 GCAGCATTTGTGATCGGCATTCATACAGACGATACGAGAGATTCGCTGCAA 1174
76 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSerTh 92
::: :::: ::||| |||::: ||||
1175 AAGAGAGAGATCAAAAGTGCACCATTCGTGCGGTGCATGATCGCTCTAC 1224
92 rProArgPylsGlnGlnAsnPheLysThrCysAlaGlnLeuValArgGlu 109
| ||||| |||::: ||||| |||::: |||
1225 CAGCGCAAGGCGGCTAATCTTTAGCCAAAGTATGAGACTGTAGTGAATAGG 1274
109 LalaIaArgLeuGluAlaCysLeuAlaPheLeuProGluAlaPheAspPhe 125
|| ::|||::: |||::: |||::: |||:::
1275 CCAAGTCACAGAACGCGCTGCATGCTCTTCTGCTAGTGTGCTGACTTT 1324
126 IleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGluGly 142
::: :::: ::|||::: |||::: |||
1325 GTGGGCGAGAGCCGAGCCCAAACTATGAGACTCTCCAGAGGCTTGGACGG 1374
142 yLysLeuLeuGlnGlnTyrThrGlnLeuAlaArgGluCysGlyLeuTrpL 159
|::: |||::: |||::: |||::: |||:::
1375 CGAGTTAATGCGCCAGTACCGGGAATTTGGCAAACTGTCACAAAAATTTTGA 1424
159 euSerLeuGluGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGln 175
::|||::: |||::: |||::: |||:::
1425 TTTCCCTGGGTGGCGTCCACAGCGGAACAT.....CAA 1459
176 LysIleTyrAsnCysHisValLeuLeuAsnSerLeuGlyAlaValAlaI 192
|||::: |||::: |||::: |||::: |||:::
1460 AAAATCTTCAACGCTCATGTTTGCTCAACGAGAAAGGGAACATACAGCC 1509
192 aThrTrpArgGlyThrHisLeuCysAspValGluIleProGlyGlnGly 209
|::: |||::: |||::: |||::: |||:::
1510 AGATATCAGAAACCTGCACCATGTTGTGATGTTCAGCACTAAA...GAGTTC 1556

```

209 rohetcysgluserasnsrThmeProgluProserleuuserPro 225  
 1557 GCGTACCGCATCGATACAGTTACCGCGGGATCTGCTTGAGCCCCA 1606  
 226 valSerThrProAlaIyLysIleGlyLeuAlaValCysTyrAspMetar 242  
 1607 GTGAGCACTTCGACTTGGCCGATAGAGGCGCTTCATGTTGGTACACCTGG 1656  
 242 gPhePProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaIleLeuT 259  
 1657 TTTTGGTGGACCGCGGCTGCTCGAGGAAGCTGGGTCGAATTTGTMA 1706  
 259 hrryrProserAlaPheGlySerIleThnGlyProAlaHisTrpGluVal 275  
 1707 CATACCATCCGCAATTCACATACGCAACCGGTAAAGCGCATGGGAATC 1756  
 276 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValAlaAlaAla 292  
 1757 CTCCTCGGGGCCAGAGCCATAGAGACTCAATGCTTTGTGGTCCCTGGCG 1806  
 292 agInCysGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetv 309  
 1807 TCAGATAGAGGTGGCACACACGACAGACAGAGTTGGGGCCACAGCATGA 1856  
 309 alyAlaSprProTrpGlyThrValAlaAlaArgCysSerGluGlyPro... 324  
 1857 TCGTTAGCCCTGCGGAACGTACTGCTGACTGCACGCGACGAGACTT 1906  
 325 GlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAlaGlnLeuArgar 341  
 1907 GATATAGGACGCGCGAGGTGACCTTCCGCTTCATTCATCTGTATCA 1956  
 341 gHisLeuProValPheGlnHisArgarProAspLeuTyr 354  
 1957 GACCATGCCCCCTGCTCGAACATGTGGAACGACATCTAC 1996  
 seq\_name: /SID1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL20222  
 seq\_documentation\_block:  
 ID ABL20222 standard; DNA; 5692 BP.  
 XX  
 AC ABL20222;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE  
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.  
 DE  
 XX Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 KW  
 XX Drosophila melanogaster.  
 OS  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-061415O.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Claim 1; SEQ ID NO 12139; 21pp + Sequence Listing; English.







CC preparation method of said protein and nucleic acid sequence, and a  
 CC method of detecting human hnit2-ase nucleic acid sequence and polypeptide  
 CC in sample.

XX Sequence 1001 BP; 264 A; 235 C; 245 G; 257 T; 0 other;

alignment\_scores:  
 Quality: 495.00 Length: 321  
 Ratio: 2.403 Gaps: 8  
 Percent Similarity: 64.174 Percent Identity: 34.891

alignment\_block:  
 US-09-357-675C-25 x AAF84214 ..

Align seg 1/1 to: AAF84214 from: 1 to: 1001

```

61 serValLeuGlyAlaGluProArgProArgAlaMetAlaIleSerSer 77
   :::::::::::::::::::: |||||:::
11 GCCATTATGGCGGGGAGGCGAGCGCGCGCTGCCGCTACCTCCAGGCG 60
   ::::: |||||:::
77 rSerCysGluLeu.ProLeuValAlaValCysGlnValThrSer..... 91
   |||::: |||||:::
61 CTCAGTCGCGCGCGAGGTGGTCTGTCTGACAGATCATGACCTCTTTC 110
   ::::: |||||:::
92 .....ThroAspLysGlnGlnAsnPh 99
   :::::::::::::::::::: |||||:::
111 CGCTGGCCCTCATCCAGCTTCAGATTCTTCCATCAATCAGATTAACGT 160
   :::::::::::::::::::: |||||:::
99 eLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysL 116
   :::::::::::::::::::: |||||:::
161 CACTCGCGCTTGTATGCTTCATCCGGAGGAGCAGCAAGCAGAGCCAA 210
   ::::: |||||:::
116 euAlaPheLeuProGluAlaPheAspPheIleAlaArgSPProAlaGlu 132
   ::::: |||||:::
211 TAGTTCTTTCGCGAATGCTT..... 233
   ::::: |||||:::
133 ThrLeuHisLeuSerGluProLeuGlyLysLeuLeuGlnGluTyr.. 148
   :::::::::::::::::::: |||||:::
234 .....AATCTCCATATGAGCGAATATTTTCTCTGAATATGCG 271
   :::::::::::::::::::: |||||:::
149 .....ThrGlnLeuAlaArg 154
   :::::::::::::::::::: |||||:::
272 AGAGAAATTCCTGCGTGAATCCACAGAACCTTCTCAATGACCAAGG 321
   :::::::::::::::::::: |||||:::
154 LucysGlyLeuThrPLeuSerLeuGlyLysPheHisGluArgGlyGlnAsp 170
   :::::::::::::::::::: |||||:::
322 AATGCACATATATCTCATTTGAGAGCTCTATCCCTGAA..... 359
   :::::::::::::::::::: |||||:::
171 TrpGlnGlnThrGlnLysIleTyrAsnGlyHisValLeuLeuAsnSerLy 187
   :::::::::::::::::::: |||||:::
360 ...GAGGATGCTGGGAAATATATATACACCTGTCTGTGTGGGCGCTGA 406
   :::::::::::::::::::: |||||:::
187 sGlyAlaValAlaThrTyrArgLysThrHisLeuGlyAspValGlu 204
   :::::::::::::::::::: |||||:::
407 TGGAACTTACTACAGAAAGTATAGAAAGATCATCTGTTTACACTGTATG 456
   :::::::::::::::::::: |||||:::
204 leProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyPro 220
   :::::::::::::::::::: |||||:::
457 TTCCTGGAATAATTCATTTCAAGAACTAAACATTGAGCGGGGTGAT 506
   :::::::::::::::::::: |||||:::
221 SerLeuLeuSerProValSerThrProAlaGlyLysIleGlyLeuAlaVal 237
   :::::::::::::::::::: |||||:::
507 AGTTTC...TCCACATTGTGATCTTACTGTCAGAGGTGGTGGCGAT 553
   :::::::::::::::::::: |||||:::
237 LcysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAla 254
   :::::::::::::::::::: |||||:::
554 CTGTAGACACTGGGTTTGCAGAGCTTGCACAAATCAGCAGCAGAGAG 603
   :::::::::::::::::::: |||||:::
254 LysAlaGluLeuLeuThrTyrProSerAlaPheGlySerIleThrGlyPro 270
   :::::::::::::::::::: |||||:::
604 GCTCCAGCTGTTGTATATCAGAGAGCTTTATCTGACACAGCAGGACCA 653

```

```

271 AlaHisTrpGluValLeuLeuArgAlaArgAlaIleGlnTrpGlnCysTyr 287
   |||||::: |||||:::
654 GCCCATTTGGAGTACTTACACGAGAACCGGCGCTGTGATATCAGGTCTA 703
   ::::: |||||:::
287 rValValAlaAlaIleGlnCysGlyArgHisGlnGlnLysArgAlaSer 304
   |||||::: |||||:::
704 TGTGGCCACAGCTCTCTCTGCC.....CGGATGACAAAGCTCTCT 744
   ::::: |||||:::
304 Yr.....GlyHisSerMetValValaAspProTrpGlyThrValVal 317
   |||||::: |||||:::
745 ATGTTCCTGGGACACACCGGTGTGAACCTTGTGGGAGGAGGTCTTA 794
   ::::: |||||:::
318 AlaArgCysSerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeuAs 334
   |||||::: |||||:::
795 GCCAAAGCTGGCAGAGAGAACATCTGTATTCAGACATAGACCTTAA 844
   ::::: |||||:::
845 GAAGCTGGCTGAATATACGCGCAAAATCCCGTTTATGACAGAAAGCAT 894
   ::::: |||||:::
351 roAspLeuTyr 354
   |||||:::
895 CAGACCTCTAT 905

```

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AA025456

seq\_documentation\_block:  
 ID AAD25456 standard; cDNA; 1359 BP.

AC AAD25456;

DT 26-MAR-2002 (first entry)

DE Human Nlt2 cDNA.

KW Human; Nlt2; cytosstatic; neuroprotective; cellular pathway; therapy;

KM apoptosis; proliferative disorder; degenerative disease; ss.

OS Homo sapiens.

FN Key Location/Qualifiers

FT CDS 22..852 /\*tag= a

FT /product= "Human Nlt2 protein"

PN W0200187958-A2.

PD 22-NOV-2001.

PF 15-MAY-2001; 2001MO-US15664.

PR 16-MAY-2000; 2000US-204713P.

PA (UYJE-) UNIV JEFFERSON THOMAS.

PI Croce C, Brenner C, Pekarski Y;

XX WPI; 2002-082984/11.

DR P-SDB; AAE15790.

PT Isolated cDNA encoding human, mouse, frog and yeast Nlt2 proteins,

PT useful to find molecules that mimic or antagonize Flt1 interaction for

PT the treatment of proliferative or degenerative diseases

PS Claim 3; Page 59; 61pp; English.

The invention relates to isolated nucleic acids comprising a fully defined cDNA nucleotide sequence encoding human, Xenopus laevis and mouse Nlt2 proteins. Nlt and Flt1 proteins are encoded as fusion proteins in invertebrates and as separate polypeptides in vertebrates. CC Nlt and Flt1 interact physically and functions in same cellular pathways. CC Molecules which bind Nlt2 and mimic or antagonise Flt1 interaction are used to treat diseases in which activity of Nlt2 protein is altered in

CC a mammal. Flt1 mimics induce apoptosis and are particularly useful to  
 CC treat proliferative disorders, whilst Flt1 antagonists promote cell  
 CC proliferation and are particularly useful to treat degenerative disease.  
 CC The present sequence is human Nit2 cDNA.

SO Sequence 1359 BP; 388 A; 284 C; 312 G; 374 T; 1 other;

#### alignment\_scores:

Quality: 485.50 Length: 286  
 Ratio: 2.542 Gaps: 6  
 Percent Similarity: 66.783 Percent Identity: 36.713

#### alignment\_block:

US-09-357-675c-25 x AAD25456 ..

Align seg 1/1 to: AAD25456 from: 1 to: 1359

```

84 ValAlaValAlcysGlnValThrSerThrProAspLysGlnGlnAsnPhely 100
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
37 TTGGCCCTCATCCAGCTTCAGATTCTTCATCAATAATCAATACATGAC 86
100 sThrCysAlaGlnValArgGlnValAlaAlaArgLysGlnValAlcysLeuA 117
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
87 TCAGCTTTAGCTTCATCCGAGGACGACGACGACGACGACGACGACGAC 136
117 lApheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaGluThr 133
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
137 TTTCCTTGGCCGGAATGCTTT..... 156
134 LeuHisLeuSerGlnProLeuGlnGlyLysLeuLeuGlnGluTyr..... 148
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
157 .....AATTCTCATATGAGGCAAAATATTTCCTGAATATGAGA 197
149 .....ThrGlnLeuAlaArgGlu 155
198 GAAATTCCTGTTGAATCCACACAGAGCTTTCTGAAGTAGCAAGCAAT 247
155 ysgLysLeuThrPheLeuSerLeuGlyLysPheHisGlnArgGlyGlnAspTyr 171
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
248 GCGCATATATTCATTCATTTGGAGGCTATCCCTGAA..... 282
172 GluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysG1 188
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
283 GAGGATGCTGGGAAATTAATTAACACCTGCTGCTGTTGGCCTGATGG 332
188 yAlaValAlaAlaThrTyrArgLysThrHisLeuCysAspValGluLeuP 205
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
333 AACTTACTAGCAAGATATAGAAAGATCCATCTGTTGACATTGATGATTC 382
205 roGlyGlnGlyPrometCysGluSerAsnSerThrMetProGlyProser 221
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
383 CTGGAAATATTCATTCATCAAGAAATCAAAATCATGAGTCCGGGTGATAGT 432
222 LeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCy 238
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
433 TTC...TCCACATTTGATACTCTTACTGACAGATGGTCTGGGCATCTG 479
238 sTyrAspMetArgPheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlyA 255
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
480 CTACGACATGCGGGTTTGCAGAGCTTCACAAATCTACGCGCAAGAGAGCT 529
255 lAluLeuLeuThrTyrProSerAlaPheGlySerIleThrGlyProAla 271
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
530 GCGAGCTGTTGATATCCAGAGCTTTTATCTGACCACTGGACCCAGCC 579
272 HisThrGluValLeuLeuArgAlaAlaGlnAlaGlnAlaGlnCysTyrA 288
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
580 CATTTGGAGTACTTCAGCGAAGCCGGCTGTGATATATCATGAGTATGAT 629
288 yAlaAlaAlaAlaGlnCysGlyArgHisGlnLysArgAlaSerTyr. 304
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
630 GCGCAGAGCTCTCTGCTGCC.....GCGGATGACAAAGCCTCCTATG 670

```

```

305 .....GlyHisSerMetValAlaAspProTrpGlyThrValAla 318
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
671 TTGGCTGGGACACAGACCGGTGTAACCTTGGGGGAGGTTCTTACGCC 720
319 ArgCysSerGlnGlyProGlyLeuCysLeuAlaArgLysLeuAsnTyr 335
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
721 AAAGCTGGCACAGAGAACCAATCGTGTATTCAGACATACCTGAAGA 770
335 lLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProA 352
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
771 GCTGCTGAATATGCGCAGCAAAATCCCGTTTTCAGACAGACGATCAG 820
352 spleuTyr 354
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
821 ACCTCTAT 828

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: AAD25457

seq\_documentation\_block:  
 ID AAD25457 standard; cDNA; 1292 BP.

AC AAD25457;

DT 26-MAR-2002 (first entry)

DE Mouse Nit2 cDNA.

KW Mouse; Nit2; cytostatic; neuroprotective; cellular pathway; therapy;  
 apoptosis; proliferative disorder; degenerative disease; ss.

OS Mus sp.

FT Key Location/Qualifiers  
 FT CDS 46..876  
 FT /tag=a  
 FT /product="Mouse Nit2 protein"

PN WO200187958-A2.

PD 22-NOV-2001.

PF 15-MAY-2001; 2001WO-US15664.

PR 16-MAY-2000; 2000US-204713P.

PA (UYE-) UNITV JEFFERSON THOMAS.

PI Croce C, Brenner C, Pekarski Y;

DR WPI: 2002-082984/11.

PS P-PSDB: AAE15791.

PT Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,  
 useful to find molecules that mimic or antagonize Flt1 interaction for  
 the treatment of proliferative or degenerative diseases

PS Claim 9; Page 59-60; 61pp; English.

CC The invention relates to isolated nucleic acids comprising a fully  
 CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and  
 CC mouse Nit2 proteins. Nit and Flt1 proteins are encoded as fusion  
 CC proteins in invertebrates and as separate polypeptides in vertebrates.  
 CC Nit and Flt1 interact physically and functions in same cellular pathways.  
 CC Molecules which bind Nit2 and mimic or antagonise Flt1 interaction are  
 CC used to treat diseases in which activity of Nit2 protein is altered in  
 CC a mammal. Flt1 mimics induce apoptosis and are particularly useful to  
 CC treat proliferative disorders, whilst Flt1 antagonists promote cell  
 CC proliferation and are particularly useful to treat degenerative disease.  
 CC The present sequence is mouse Nit2 cDNA.

SO Sequence 1292 BP; 354 A; 292 C; 289 G; 357 T; 0 other;

alignment\_scores:                   Quality: 481.00                   Length: 302  
                                   Ratio: 2.324                   Gaps: 9  
                                   Percent Similarity: 68.543           Percent Identity: 37.748

alignment\_block:  
 US-09-357-675c-25 x AAD25457 ..

Align seg 1/1 to: AAD25457 from: 1 to: 1292

```

59 GlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSe 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7 CAGGATCCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 48
75 rSerSerCysGlnLeuProLeuValAlaValCysGlnValThrSert 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49 ....TCTACTTCCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 91
92 hProAspLysGlnGlnAsnPhelYstHrcysAlaGluLeuValArgLu 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 TT.....AAATCAGATACCTTACCCTGGCTGGCTGGCTGGCTGG 135
109 AlaAlaArgLeuGlyAlaCysLeuAlaPhelProGlnAlaPhaAsp 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GCGAGCAAGACAGAGTCCCAACATAGTTCTCTGCTGAGTGTTCAT. 183
125 eIleAlaArgAspProAlaGluThrLeuHisLeu.....SerGluP 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....TCTCATATGAGAACACCTTCTCTGCTGGCTGGCTGGCTGG 223
139 rIleuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGlu 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 AGATTCTCTGGAGAGTCCACACAAAGCTTCTGAGATGACAAAGAG 273
156 GlyLeuThrLeuSerLeuGlyGlyPheHisGluArgGlyGlnAsp 172
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 AGCATATATCTCATGTGAGGCTCCATCCCTGAA.....GA 308
172 uGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLys 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 GGAATGCTGGGAAACTGTATATACCTGCTGTTGGCTGGCTGGATG 358
189 IAlaValAlaThrTyrArgLysThrHisLeuCysAspValGluIle 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 GTTTACTGTGAAAGACACAGAGATCATCTGTTGACATGTGATGCT 408
206 GlyGlnGlyPrometCysGlnSerAsnSerThrMetProGlyProse 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
409 GGGAAATTAATCGTTTCAGAGATCTAAACATTGAGCCCTGGTATG 458
222 uGlnSerProValSerThrProAlaGlyLysIleGlyLeuAlaVal 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
459 C...TCCACATTTGATACGCCCTTACGCAAAAGGCGCTGGCATG 505
239 YrAspMetArgPheProGlnLeuSerLeuAlaLeuAlaGlnAlaG 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
506 ATGATATGCGCTTCGGGACCTTGCAACAAATCATGACCAAAAGG 555
256 GlnIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAl 272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
556 CAGCTCTGGTGTATCTGGAGCTTTCAATCTGACCCACAGACCA 605
272 sTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyr 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
606 CTGGAGCTCTCTCAGGAGCCCGGCTGTGATATCAGATGATGTGG 655
289 aIleAlaIleAlaGlnCysGlyArgHisHisGluLysArgAlaSer 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
656 CTACAGGCTCTCTGCT.....CGGATGACAAAGCCTCATGTG 696
305 .....GlyHisSerMetValAlaAspProTyrProGlyThrValAl 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

697 GCCTGGGACACAGCACTGTTGTGATCTTGGGGCAGGCTTAACCA 746
319 gCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsn 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
747 AGCTGGCAGCAGAGGAAACAAATCTGTACTCAGATCATGACCTGA 796
336 eAlaArgGlnLeuArgArgHisLeuProValPheGlnHisArgArg 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
797 TGCTGAATTTGGGACGACAAATCCCATTTTAAACAGAAAGCAGCA 846
353 LeuTyr 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
847 CTCTAT 852
seq_name: /STDS1/gcgdata/hold-geneseq/geneseqn-emb1/AA81479
seq_documentation_block:
ID AA81479 standard; DNA; 69936 BP.
XX
AC AA81479;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
XX
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V,
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 547-567; 1760bp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AA81453 to AA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AA81260 to AA81303 and AA825620 to AA825663 represent
CC Neisseria DNA sequences and their corresponding proteins; AA81254 to
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AA81322 to
CC AA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.

```



CC AAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.

XX Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

alignment\_scores:  
 Quality: 404.00 Length: 307  
 Ratio: 2.138 Gaps: 8  
 Percent Similarity: 61.564 Percent Identity: 35.179

alignment\_block:

US-09-357-675C-25 x AAF21607 ..

Align seg 1/1 to: AAF21607 from: 1 to: 349980

51 LeucysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67  
 |||||  
 153960 TTGTGTAAA.....CGTTTTCGCAAGTACCTCATGTGATGCGTATCG 154003  
 67 OArg.....ProArgAlaMetAlaIleSerSerSerSerCysLeuPr 82  
 |||||  
 154004 GCGGAAATATGCCGCAAGGTCGCGCAAGAAAGAGAAAGAAATGCAAA 154053  
 82 roleuValAlaValAlaCysGlnValThrSerThrProAspLysGlnGlnAsn 98  
 |||||  
 154054 TCAGAGTTGGCGCGCGTCAGATGTCGCGCGCGTGTGCGCGCAAAACAC 154103  
 99 PheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCys 115  
 |||||  
 154104 GTCGCCCGCATGAAGCCGTCGTGCGCACGCGCGCGAGCGGTGCGGA 154153  
 115 sleuAlaPheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaG 132  
 |||||  
 154154 TTGGGTCTGTGCGCGCAATATGTCGTGATGTCGCGCAACGATACCG 154203  
 132 IuThrLeuHisLeuSerGluProLeu...GlyGlyLysLeuLeuGlu 147  
 |||||  
 154204 ACAAACTCGGCGCTTGGCGAGCCTTGGGCGGCGGACGCTTCACAGCGCA 154253  
 148 TyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 163  
 |||||  
 154254 TTGAGCGAAACGCGAAAGATGCGCGGTGCTG...TTCGCGCGGAC 154300  
 164 .....PheHisGluArgGlyGlnAspTrpGluInThrGlnL 176  
 |||||  
 154301 TGTGCCCTGCAGAACTGTCGAGCGCGGT.....A 154329  
 176 ysiLeuTrpAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 192  
 |||||  
 154330 AAGTGAATGAAATACGCTGTGTGTCAGCAGCGAGCGCGTGAAGACGCGG 154379  
 193 ThrTyArgLysThrHisLeuCysAspValAluIleProGlyGlnGlyPr 209  
 |||||  
 154380 CTGTACACCAAAATGACACTC.....TTCGGTTTTCGCTTGGCGGA 154423  
 209 oMetCysGluSerAsnSerThrMetProGlyProSerLeuLeuSerProv 226  
 |||||  
 154424 ACGCTATGCCGAGCGATACCATCCGCGGCGGCGGATGTGCCGCACT 154473  
 226 alSerThrProAlaGlyLysIleGlyLeuAlaValAlaCysTyArgMetArg 242  
 |||||

154474 TGTGCGCAGAAAGCGCGTGGCGGCGGCGCATTTGTTACGATGTCGCC 154523  
 243 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 259  
 |||||  
 154524 TTTCGCCGAA.....TTTTTCGACGCGCATGTCGCTTGTGACGATTTAT 154567  
 259 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValL 276  
 |||||  
 154568 GCTGCCCGCTGCGTTTACGACACGACGCGGCGCAAGCGCATTTGGAGCTGC 154617  
 276 euleuArgAlaArgAlaIleGluThrGlnCysTyValAlaAlaAla 292  
 |||||  
 154618 TGTGCGCGCGCGTGGCGTGCAGAAACCAATGTATGCTGCGTGGCGCGCA 154667  
 293 GlnCysGlyArgHisHisGluLysArgAlaSerTyArgLysSerMetVal 309  
 |||||  
 154668 CAGGCGGCTTTCGACGAAACGAGCGGCGCGCATGTCGCGACACGATAT 154717  
 309 lValAspProTrpGlyThrValAlaAlaArgCysSerGluGlyProGlyL 326  
 |||||  
 154718 TGTGATCTCGTGGCGCGACGTCGTTGTGACGTATTCGCCGAGGCGGAAGCG 154767  
 326 eucysLeuAlaArgIleAspLeuAsnTyTrpLeuArgGlnLeuArgHis 342  
 |||||  
 154768 TTGTTACGCGACAGATCGATGCGCAACCGCTGAAACAGCGTCCGCAACCGC 154817  
 343 LeuProValPheGlnHisArg 349  
 |||||  
 154818 CTGCCCGCTTGAAATACCGG 154838  
 seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA81490  
 seq\_documentation\_block:  
 ID AAA81490 standard; DNA; 1437668 BP.  
 XX  
 AC AAA81490;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KM Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN MO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-0523573.  
 XX  
 PR 09-OCT-1998; 98US-0103794.  
 XX  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Maignant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;  
 PI Rappuoli R, Pizza M;  
 XX  
 DR WPI; 2000-318079/27.  
 XX  
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -  
 XX  
 PS Claim 7; Page 866-1272; 1760pp; English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA

sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA sequences and their corresponding primers; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition, and the composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes; and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

XX Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

alignment\_scores:  
Quality: 404.00 Length: 307  
Ratio: 2.138 Gaps: 8  
Percent Similarity: 61.564 Percent Identity: 35.179

alignment\_block:  
US-09-357-675C-25 x AAA81490 ..

Align seg 1/1 to: AAA81490 from: 1 to: 1437668

```

51 LeucylProglyLeuArgIleProGlnLeuSerValLeuCysAlaGlnP 67
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
453960 TTGTGTAAA.....CGTTTTCGCAAGATACCGTCATGTCGTCG 454003
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 GAG.....ProArgAlaMetAlaIleSerSerSerSerCysGluLeuP 82
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454004 GCGGAATATGCGCCAGAGTCGCGCAGCAAGAAAGAAATGACAAA 454053
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 roLeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsn 98
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454054 TCAGAGTTCGCGCGTCAGATGTGTGCGGCGTTCGCGCGCAAC 454103
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 PheLysThrCysAlaGlnLeuValArgGluAlaAlaArgLeuGlyAla 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454104 GTGCGCGCATGAAACGCGTGTGCGCGCGCGCGAGCAGGAGTTCG 454153
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 IleuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAla 132
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454154 TTGGGTGCTGCTCCCGAATATTGGGTGCTGATGGCGCAAGATACG 454203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 IuThrLeuHisLeuSerGluProLeu...GlyLysLeuLeuGluGlu 147
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454204 ACAAACTCGGCTTTCGCGAGCCTTTCGCGCGCGAGCCTTTCAG 454253
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 TyrThrGlnLeuAlaArgLysGlyLeuThrLeuSerLeuGly.. 163
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454254 TTGAGCGAAACGCGCAAGATGCGCGTGTGCTG...TTTCGCGG 454300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 .....PheHisGluArgGlyGlnAspTyrPglGlnThrGlnL 176
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454301 TTGCGCGCTGCAAGCTCGAGCGGCT.....A 454329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
176 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 192
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454330 AAGTGTGAATACGCTGTGTGTCAGGACGCGAGCGCTGAAGACGG 454379
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
193 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGly 209

```

```

454380 CTGTACCAAAAATGACACCTC.....TTGGGTTTTCGCGTTTGCGCA 454423
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
209 omecYsGluSerAsnSerThrMetProGlyProSerLeuGlySerPro 226
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454424 ACCGTATGCCGAAGCCGATACATCCGCGCGCGGAGTGTCCGACT 454473
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 alSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArg 242
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454474 TTGCGCAGAAAGCGTCCGTCGCGCGCGGCAATTGTTCAGATGTC 454523
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
243 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluLeuThr 259
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454524 TTTCGCCAA.....TTTTCGACGCGCAGTCCGCTTTCAGCTATTGAT 454567
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
259 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTyrGluVal 276
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454568 GCTGCCGCTGCTGTTACGACAGCGGCAAGCCGATTTGGAGCTGC 454617
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 euleuArgAlaArgAlaIleGluThrGlnCysTyrValAlaAlaAla 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454618 TGCTGCGCGCGCGTCCGTCGCAAAACCAATGTTACGTCGCGCGCGCA 454667
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
293 GlnCysGlyArgHisGlnLysArgAlaSerTyrCylHisSerMetVal 309
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454668 CAGGCGGCTTTCGACGCAAGAGCGCGCAGTTCGACACAGCATGAT 454717
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
309 lValAspProTyrGlyThrValAlaAlaArgCysSerGlnGlyProGly 326
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454718 TTGTCATCCCTGGGCGCGACGCTTTGGACGTAATTCGCCGAGCGCA 454767
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
326 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 342
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454768 TTGTTACGCGCAGACATGATGATCCACCGCTGAACAGCGTCGCA 454817
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
343 LeuProValPheGlnHisArg 349
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454818 CTGCGCGCTTGAATACCGG 454838
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq\_name: /SIDSI/gcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC77176

seq\_documentation\_block:

ID AAC77176 standard; cDNA; 1725 BP.

```

XX AC AAC77176;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF2731 polynucleotide sequence SEQ ID NO:5461.
XX KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
XX KW vulnerability; antiproliferative; antiparkinsonian; neuroprotective;
XX KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
XX KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX KW hypotensive; dermatological; immunosuppressive; antinflammatory;
XX KW antiviral; antibacterial; antifungal; antineuritic; antithyroid;
XX KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX KW cholesterol ester storage; systemic lupus erythematosus; infection;
XX KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX KW bone damage; cartilage damage; antinflammatory disease; coagulation;
XX KW thrombosis; contraceptive; ss.
XX OS Homo sapiens.
XX PN WO200058473-A2.
XX PD 05-OCT-2000.
XX PF 31-MAR-2000; 2000WO-US08621.
XX

```







```

312 ProTrpGlyThrValAlaAlaArgCysSerGluGlyProGlyLeuCysLe 328
|||||::: ||| :::| ||||| ||| :::
746 CCATGGGCCCAAGCTGCAGACAGTGCCTGAAGCGCAGGAATCGTGT 795
328 uAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuPro 345
::||| |||||::: ::: |||||::: |||||
796 GCCCGATATGATTTCTCCGAGTGGAGACAGTGCCTCAGCAGATCCCG 845
345 aPheGlnHisArgArgProAspLeuTyr 354
||||| ::: ||||| |||||
846 TCTTTGGGCAAGACGCTCTAGATCTGTAC 874

```

THIS PAGE BLANK (USPTO)



Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,  
URL: <http://genome-gsc.riken.go.jp/>, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
Please visit our web site (<http://genome-gsc.riken.go.jp/>) for  
further details.

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5'-GAGGAGAGAGCGCGCCACCTCGACTGTTTCTTTTCTTNN 3'], cDNA was  
prepared by using triethanolamine-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot = 5.0. Second strand cDNA  
was prepared with the primer adapter of sequence [5'-  
GAGAGAGAGAGATCCAGATCAATTAATTAATTAACCCCGCCCG 3']. cDNA was  
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
XhoI. Host: SOLR.

FEATURES  
source Location/Qualifiers

1..1374  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="MGI:1891931"  
/db\_xref="taxon:10090"  
/clone="061007B07"  
/sex="male"  
/tissue\_type="kidney"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
1..1374  
/gene="Ntli"  
1..1374  
/note="data source:MGI, source key:MGI:1350916,  
evidence:ISS  
ntrlase 1"

BASE COUNT 330 a 359 c 342 g 343 t  
ORIGIN

alignment\_scores:  
Quality: 1529.00 Length: 364  
Ratio: 4.662 Gaps: 4  
Percent Similarity: 90.110 Percent Identity: 81.044

alignment\_block:  
US-09-357-675c-25 x AK002269 ..

Align seg 1/1 to: AK002269 from: 1 to: 1374

1 ProLeuAlaAlaAla\*\*LeuAlaProAspArgProAspArgThrLe 17  
||||||| ||| ||||||| ||| |||  
10 CCGCTGCTGCTGCGCTGCTGCTGCGAGCTTGAGATCGACCCC 59  
17 uArgMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 34  
||||||| ::||| ||||||| ||| ::|  
60 GTGAATGTTTGGCTGCTGCTCATGTAGAGCTACTCCCAACCTGC 109  
34 rArgPro.ArgLeuGlyPheIleThrArgProProHisArgPheLeuSerLe 50  
||||||| ||||||| ||||||| ||| ::|  
110 CCGCTGCGGCTGCTCATCACAGGCTCTCCACCA.....CT 150  
50 uLeuCyProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnP 67  
||||||| ||| ::| ||||||| ||| ::|  
151 CCGTGTACCGGATACCGATTACTCGAACCCAGTACTTTGTACTCAGC 200  
67 rArgProArgAlaMetAlaIleSerSerSerCysGlnLeuProLeu 83  
||||||| ::||| ||||||| ||| |||||||  
201 CCAAGGCCCAACCATGTCC...TCATCAACTTCTCGAGAGCTCCCTG 247  
84 ValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPhely 100  
||||||| ::||| ||||||| ||| ::|||  
248 GTGGCTGTGTGCGAGTACATCAACCAACCAAGCAAGAACTTAA 297

100 sThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeu 117  
||||||| ::||| ||||||| ||| |||||||  
298 AACATGTCTAGTGGTTGACAGAGGCTGCCAGACTGGTCTGCTGCTGG 347  
117 LaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThr 133  
||||||| ::||| ||||||| ||| ::|||  
348 CTTTCTGCTGAGGAGCTTTCATCTTATTCGACCAACCTCGCCGAGACA 397  
134 LeuHisLeuSerGluProLeuGlyGlyLeuLeuGluGluTyrThrGln 150  
||| ||||||| ||||||| ||| ||| ::|||  
398 TTACTCTGTCCGACCACTGAATGGGATCTTTGGGCCAATATACCA 447  
150 nLeuAlaArgGluCys.GlyLeuThrLeuSerLeuGlyGlyPheHisGlu 166  
||||||| ||| ::||| ||||||| ||| |||||||  
448 GCTTCCAGGGAATTTGGAATCTGGCTGCTGGCGGTTCCACAGAG 497  
167 ArgGlyGlnAspTrpGluGlnThrGlnLysIleThrAsnCysHisValLe 183  
||||||| ::||| ||||||| ||| ::|||  
498 CTTGCCCAAGACTGGAGCAGAAATCAAAATCTACATTCATGTGCT 547  
183 uLeuAsnSerLysGlyAlaValAlaIleThrArgLysThrHisLeu 200  
||||||| ||| ::||| ||||||| ||| |||||||  
548 TTGAACAGCAGAGGATCAGTAGTGGCCAGTTACGAGAGACACATCTGT 597  
200 yAspValGluIleProGlyGlnGlyPrometCysGluSerAsnSerThr 216  
||||||| ||| ::||| ||||||| ||| |||||||  
598 CGGATGTAGAGATCCAGCTCAGGGCCGATGACAGAAACACATATACC 647  
217 MetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 233  
||||||| ::||| ||||||| ||| ::|||  
648 AAGCTGAGGACACTTGGAGCCACTGTCAAGACACCGGCTGCCAAGT 697  
233 eGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAla 250  
::||| ||||||| ::||| ||||||| ||| |||||||  
698 TGGCTAGCAATCTGTATGACATGCGGTTCCGAACTTCTTGAAT 747  
250 eAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySer 266  
||||||| ||| ::||| ||||||| ||| |||||||  
748 TGGCTCAAGCTGGGGCAGAAATACTTACTTATCTTCAGCCTTACATCT 797  
267 IleThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleG 283  
::||| ||||||| ::||| ||||||| ||| ::|||  
798 GTTACAGCTCCAGCCCTGAGAGTGTCTGTGGGCGCCGCGCATTTGA 847  
283 uThrGlnCysTyrValAlaAlaAlaGlnCysGlyArgHisGluL 300  
::||| ||||||| ::||| ||||||| ||| |||||||  
848 ATCTCAGTGTATGTATATACGACGCGCAGTGTGGACGCCACCATGAA 897  
300 yArgAlaSerTyrGlyHisSerMetValAlaAspProTrpGlyThrVal 316  
||||||| ||| ::||| ||||||| ||| |||||||  
898 CAAAGCAGCAATTATGGCCATAGATGTGGTGTGACCCGTGGGCGACAGT 947  
317 ValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLe 333  
||||||| ::||| ||||||| ||| ::|||  
948 GTGGCCCGCTGCTCCGAGGAGCAGGCGCTGCTTCTCGAATTTGATCT 997  
333 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValIlePheGlnHisArg 350  
||| ::||| ||| ::||| ||||||| ||| |||  
998 CCACTTCTCAACACAGATGCGCCAAACACTGCTGTCTTCAGACCCGCA 1047  
350 rArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 362  
||||||| ||| ::||| ||||||| ||| ::|||  
1048 GACCTGACCTCTATGAGCAGTGTGGGTATCTACACTCTCT 1085

seq\_name: gb\_hlc:AK004988

seq\_documentation\_block:

LOCUS AK004988 2993 bp mRNA linear HTC 19-JAN-2002  
DEFINITION Mus musculus adult male liver cDNA, Riken full-length enriched  
library, clone:1300013f05:ntrlase 1, full insert sequence.  
ACCESSION AK004988  
VERSION AK004988.1 GI:12836590



```

113 yalaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsp 130
|||||
291 TGCTTGCTGAGCTTCTTCCGTCGAGCATTTGACTTATTGACGAAACC 340
|||||
130 roAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGlu 146
|||||
341 CTGGCGGAGCATTAATCTCTCCGACCACTGAATGGGATCTTTTGGCC 390
|||||
147 GluThrGlnLeuAlaArgGluCysGlyLeuThrPheLeuSerLeuGly 163
|||||
391 CAATATAGCCAGCTTCCGAGGAAATGGAATCTGCTGTCTTGGCGG 440
|||||
163 yPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsn 180
|||||
441 TTTCACAGAGGTGGCCAGACTGGAGCAGAAATCAGAAATATCAAT 490
|||||
180 yHisValLeuLeuAsnSerLysGlyAlaValAlaIleThrTyrArgLys 196
|||||
491 GTCATGTCCTTTGACAGCAGAGGATCAGTGTGGCCAGTACAGGAG 540
|||||
197 ThrHisLeuCysAspValGluIleProGlyGlnGlyPrometCysGlu 213
|||||
541 ACACATCTGTGCGATGTAGATCCAGGTCAGGCGCCGATGAGAGAA 590
|||||
213 rAsnSerThrMetProGlyProSerLeuGluSerProValSerThrPro 230
|||||
591 CAATATACCAAGCCCTGAGGAGCACTTGAAGCCACTGTCAAGACCC 640
|||||
230 laGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGlu 246
|||||
641 CTGGCAAGGTGTGTAGCAATCTGTATGACATGCGGTTCCCGAACT 690
|||||
247 SerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSer 263
|||||
691 TCTTTGAATTTGCTCAAGCTGGGCGAGAAATCTTACTTATCTTGA 740
|||||
263 apHeGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAla 280
|||||
741 CTTTGATCTGTGTACAGTCCAGCCCACTGGAGAGTGTGCTGGGCG 790
|||||
280 rgAlaIleGluThrGlnCysTyrValAlaAlaAlaGlnCysGlyArg 296
|||||
791 GCGGCATTTGATCTCAGTGTATGATAGCAGCAGCAGTGTGAGCG 840
|||||
297 HisHisGluLysArgAlaSerTyrGlyHisSerMetValAlaAspPro 313
|||||
841 CACCATTAACAAGACAGATGATGCGCATGATGCTGTGACCCGCTG 890
|||||
313 rgLysThrValAlaAlaArgCysSerGluGlyProGlyLeuCysLeuAla 330
|||||
891 GGGCACAGTGTGGCCGCTGCTCCAGGAGCAGGCTCTGCTGCTGCT 940
|||||
330 rgLysAspLeuAsnTyrLeuArgGlnLeuArgAlaHisLeuProValPhe 346
|||||
941 GAATGTATCTCCACTTTCTACAAAGATGGCGCCAAACACCTGCTG 990
|||||
347 GluHisArgArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 362
|||||
991 CAGCAGCGCAGACCTGACCTTATGACAGTCTGGGTATCATCCTCT 1038
|||||
seq_name: gb_est1:AL520768
seq_documentation_block:
LOCUS AL520768 890 bp mRNA linear EST 13-FEB-2001
DEFINITION AL520768 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 5
prime, mRNA sequence.
ACCESSION AL520768
VERSION AL520768.1 GI:12784261
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 890)
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source 1. 890
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB002YD05"
/clone_1ib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 191 a 251 c 241 g 206 t 1 others
ORIGIN
alignment_scores:
Quality: 1524.00 Length: 306
Ratio: 5.201 Gaps: 1
Percent Similarity: 95.752 Percent Identity: 95.752
alignment_block:
US-09-357-675C-25 x AL520768 ..
Align seg 1/1 to: AL520768 from: 1 to: 890
7 LeuAlaProAspArgProProAspArgThrLeuArgMetValLeuAlaI 23
|||||
4 CTGGCTCCAGACCCCTCCGAGATCGACCTGCGAATGTTTGGCTAT 53
|||||
23 eSerSerCysArgThrTyrSerLeuSerArgArgProArgLeuGlyPheI 40
|||||
54 ATCTTCATGCTG.....GGCTCA 72
|||||
40 leThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLeuArg 56
|||||
73 TCACACAGCTCTCCACAGATTCCTGCTGCTGTGCTGCTGAGCTCGG 122
|||||
57 IleProGlnLeuSerValLeuCysAlaGlnProArgProAlaMetAl 73
|||||
123 ATACCTCAACTCTCAGTACTTGTGTGCTCAGCCCAAGGCCAGAGCCAT 172
|||||
73 alleSerSerSerSerCysGluLeuProLeuValAlaValCysGlnVal 90
|||||
173 TATCTCTTCTCTCTCTGCGAAGTGCCTGTGGTGTGGCCAGGTAA 222
|||||
90 hrSerThrProAspLysGlnGlnAsnPheLysThrCysAlaGluLeuVal 106
|||||
223 CATTCAGGCCAGACACAAACAGAACTTTAAACATGTGCTGAGCTGTT 272
|||||
107 ArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAla 123
|||||
273 CGAGAGGCTGCCAGACTGTGGTGTGCTGCTGCTGCTGCTGAGCAT 322
|||||
123 eaSPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluPro 140
|||||
323 TGACTTCATTGACAGGAGCCCTGACAGAGCATTACACTGTGTGAACAC 372
|||||

```

```

140 euglyglyleuLeuGlulGlyThrGlnLeuAlaArgGlyCysGly 156
|||||
373 TGGGTGGGAACCTTTGGAGAATACACCCAGCTTCCAGGAATGTGGA 422
|||||
157 LeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGlu 173
|||||
423 CTCTGGCTGCTGGGTGGTTCATGACGATGCGCCAGACTGGAGCA 472
|||||
173 nThrGlnLysIleTyraAsnGlyHisValLeuLeuAsnSerLysGlyAla 190
|||||
473 GACTCAGAAATCTACAATTGTCACTGCTGTAACAGCAAGAGGGCAG 522
|||||
190 aValAlaThrTyraArgLysThrHisLeuGlyAspValGluIleProGly 206
|||||
523 TAAGTGGCACTTACAGAGACACATCTGTGTGACGTAGAGATTCCAGGG 572
|||||
207 GlnGlyProMetCysGluSerAsnSerThrMetProGlyProSerLeuG 223
|||||
573 CAGGGGCTATGTGTGAACCACTACCATGCTGGCCCACTCTGA 622
|||||
223 uSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCysTyra 240
|||||
623 GTCACTCTGACACACACAGCAGCAAGATTGCTAGCTGTCTGTATG 672
|||||
240 spMetArgPheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlu 256
|||||
673 ACATGCGCTTCCCTGACACTCTCTCTGCAATGGCTCAAGCTGGACAG 722
|||||
257 IleLeuThrTyProSerAlaPheGlySerIleThrGlyProAlaHisTr 273
|||||
723 ATACTTACCTATCCTTCAGCTTTGGATTCATTCAGGCCAGCCCACTG 772
|||||
273 pGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyraValAla 290
|||||
773 GGAGGTGTCTGTGGGCGCTGTATCGAAACCCAGCTGCTATGATGG 822
|||||
290 IaIaIaGlnCysGlyArgHisHisGlnLysArgAlaSerTyrglyHis 306
|||||
823 CAGCAGACAGTGTGGACGCCACCATGAGAGAGCAAGTTATGGCCAC 872
|||||
307 SerMetValValAspPro 312
|||||
873 AGCATGGTGGKAGACCC 890
|||||
seq_name: gb_est1:AL529152
seq_documentation_block:
LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001
DEFINITION AL529152 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSDD004YM06 5
prime, mRNA sequence.
ACCESSION AL529152
VERSION AL529152.1 GI:12792645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 844)
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 844
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDD004YM06"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"

```

```

/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by life technologies. Contact: Feng Liang life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 182 a 238 c 235 g 189 t
ORIGIN
alignment_scores:
Quality: 1413.00 Length: 263
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-357-675C-25 x AL529152 ..
Align seg 1/1 to: AL529152 from: 1 to: 844
100 LysThrCysAlaGluLeuValArgGlnAlaAlaArgLeuGlyAlaCysLe 116
|||||
1 AAAACATGTGCTGAGCTGGTTCGAGAGGCTCCACAGCTGGCTGCTGCT 50
|||||
116 uAlaPheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaGly 133
|||||
51 GCGTTTCTGCTGAGCATTTGACTTTCATTGACAGGAGACCCTGCAGAGA 100
|||||
133 hIleuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGlyThr 149
|||||
101 CGCTACACCTGCTGACACCACTGGGTGGGAAACTTTTGAAGAATACACC 150
|||||
150 GlnLeuAlaArgGlnCysGlyLeuThrLeuSerLeuGlyGlyPheHisG 166
|||||
151 CAGCTTGCAGGGAATGTGACTGTGCTGCTGGTGGGTGGTTCATGA 200
|||||
166 uArgGlyGlnAspTrpGluGlnThrGlnLysIleTyraAsnGlyVal 183
|||||
201 GCGTGGCCAGACTGGAGACAGACTCAGAAATCTACAAATTGTCTACGTC 250
|||||
183 euleuAsnSerLysGlyAlaValAlaIaThrTyraGlyThrHisLeu 199
|||||
251 TGCTGACAGCAAGAGGCGAGTGGCCACTTACAGAGAACACATCTG 300
|||||
200 CysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerTh 216
|||||
301 TGTGACGTAGAGATTCCAGGGCAGGGGCTGTATGTGAACCACTCTAC 350
|||||
216 rMetProGlyProSerLeuGlnSerProValSerThrProAlaGlyLys 233
|||||
351 CATGCTTGGCCCAAGTCTTGAAGTCACTGTCAAGCACACAGCAGCAAGA 400
|||||
233 leGlyLeuAlaValCysTyraAspMetArgPheProGluLeuSerLeuAla 249
|||||
401 TTGGCTAGTGTGTGCTGATGACATGCGGTTCCCGTCAAGCTCTCTGCA 450
|||||
250 LeuAlaGlnAlaGlyAlaGluIleLeuThrTyProSerAlaPheGlySe 266
|||||
451 TTGGCTCAAGCTGGAGCAGAGATACTTTCATCTTACGCTTTTGGATC 500
|||||
266 rIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaIleG 283
|||||
501 CATTCAGAGCCCAAGCCCACTGGAGAGTGTGCTGGGCCCCCTGCTATCG 550
|||||
283 luthrGlnCysTyraValAlaAlaAlaGlnCysGlyArgHisHisGlu 299
|||||
551 AAACCACTGCTATGTAAGTGGCAGCAGCAGTGTGAGCCACCATGAG 600
|||||

```

300 LysArgAlaSerTyrGlyHisSerMetValAlaAspProTrrpGlyThrVa 316  
|||||  
601 AAGAGAGCAAGTTATGGCCACAGATGTGTAGACCCCTGGGGAAACAGT 650  
|||||  
316 ValAlaIleArgCysSerGluGlyProGlyLeuCysLeuAlaIleAspL 333  
|||||  
651 GGTGGCCCGCTGCTGAGGGGGGAGGCCCTGCTGGCCGGAATAGACC 700  
|||||  
333 eusAnfYrLeuArgGlnLeuArgAlaArgHisLeuProValPheGlnHisArg 349  
|||||  
701 TCACACTATCTGCACACAGTTGCGCCGACACCTGCTGTGTTCCAGACCCG 750  
|||||  
350 ArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 362  
|||||  
751 AGGCGTACCTCTATGGCAATCTGGGTACACCACTGCT 789  
seq\_name: gb\_est2:B1822844

seq\_documentation\_block:  
LOCUS B1822844 928 bp mRNA linear EST 04-OCT-2001  
DEFINITION 603040166F1 NIH\_MGC\_115 Homo sapiens cdna clone IMAGE:5180897 5',  
mRNA sequence.  
ACCESSION B1822844 GI:15934394  
VERSION B1822844.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 928)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LAM11451 row: d column: 18  
High quality sequence stop: 822.  
Location/Qualifiers  
source 1..928  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5180897"  
/clone\_11b="NIH\_MGC\_115"  
/lab\_host="DH10B"  
/note="organ: pooled brain, lung, testis; Vector:  
pCMV-Sport6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA  
source anonymous pool of 6 male brains, age range 23-27; 1  
male lung, age 27; and 1 male testis, age 69. Library is  
oligo-dt primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.8 kb,  
insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by C.  
Gruber (Invitrogen). Research Genetics tracking code  
021. Note: this is a NIH-MGC Library."  
BASE COUNT 198 a 256 c 259 g 214 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 1390.50 Length: 318  
Ratio: 4.762 Gaps: 5  
Percent Similarity: 91.824 Percent Identity: 90.566

alignment\_block:  
US-09-357-675c-25 x B1822844

Align seg 1/1 to: B1822844 from: 1 to: 928

13 ProAspArgThrLeuArgMetValIleuAlaIleSerSerCysArgThrTy 29  
|||||  
1 CCGGATCGACCTCGCAATGATGTTTGGCTATATCTTCAATGCTG..... 44  
29 rSerLeuSerArgArgProArgLeuGlyPheIleThrArgProProHisA 46  
|||||  
45 .....GGCTTCATCACCAGCCCTCCTCACA 69  
46 rPheLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 62  
|||||  
70 GATTCCTGTCCTTCTGTCTGTCTGAGACTCCGATACCTCAACCTCAGTA 119  
63 LeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCy 79  
|||||  
120 CTTGTGCTCAGCCGACGAGCCAGGACCATGCTATCTCTCTTCTCTCTG 169  
79 sGluLeuProLeuValAlaValCysGlnValThrSerThrProAspLysG 96  
|||||  
170 CGAAGTGCCTGCTGTGCTGTGCTGAGGCAATGACTTCAATGACGGGA 219  
96 InGlnAsnPhelysThrCysAlaGluLeuValArgGluAlaIleArgLeu 112  
|||||  
220 AACAGAACTTTAAACATGTGCTGAGCTGCTGAGAGCTGCCAGACTG 269  
113 GAlaAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAs 129  
|||||  
270 GGTGCTGCTGCTGCTTCTGCTGAGGCAATGACTTCAATGACGGGA 319  
129 pProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuG 146  
|||||  
320 CCTGCAGAGACGCTACACCTGTCTGAAACACACGCGGGAACCTTTGG 369  
146 InGluTyrThrGlnLeuAlaArgGluCysGlyLeuThrPleuSerLeuGly 162  
|||||  
370 AAGAATATACACCAGCTTGGCCAGGAAATGTGACTGCTGCTCTGGGT 419  
163 GAlaPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 179  
|||||  
420 GGTTCATGAGCGGTGCGCCAGACTGGGACGACAGCTCAGAAATCTRCA 469  
179 nCysHisValLeuLeuAsnSerLysGlyAlaValAlaIleThrTyrArgL 196  
|||||  
470 TTGTCAAGTCTGCTGTGACAGCAAGGGCAGTAGGGCCACTTACAGGA 519  
196 yThrHisLeuCysAspValGluIleProGlyGlnGlyPrometCysGlu 212  
|||||  
520 AGACACATCTGTGTGACGTAGAGATTCCAGGCAG. GGGCCTATGTGGAA 568  
213 SerAsnSerThrMetProGlyProSerLeuGlnSerProValSerThrPr 229  
|||||  
569 AGAAGCTTACATGCTGCTGGCCAGCTTGTAGTACACTTACAGCCACC 618  
229 oAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGluL 246  
|||||  
619 AGAGAGCAATGCTGTCTACTGCTGTGCTATGACATCGGTTCCCGAAG 668  
246 eusSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThr. TyrProse 262  
|||||  
669 TCTCTCTGGCATTTGGCTCAAGCTGAGCAGAGATTAATTACTTATCTTC 718  
262 rAlaPheGlySer...IleThrGlyProAlaHisTrpGluValLeuLeuA 278  
|||||  
719 AGCTTTTGGATCCCATATACAGGCGCCAGCCACTGGAGGCTGTGCTGC 768  
278 rGAlaArgAlaIleGluThrGlnCysTyrVal. ValAlaIleAlaGlnCy 294  
|||||  
769 GGGNCCGTGCTATCGAAACCCAGTGTATGTAGTGGCGGCACACAGCTG 818  
294 sGly. ArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValVal 310  
|||||  
819 TGAAGCGCCACCATGAGAGAGAGAGC. AGTTATGGGCACACAGATGTGTGA 867



311 AspProtpcglyThrValAlaAlaArgCysSerclncljProglyLeu 326  
 :::||| ||| ::| |||||  
 868 AACCTGGGGACAGGGGCCCTGCTCTGAGGGAGCCCTGGCTTG 915

seq\_name: gb\_est2:BI769604

seq documentation block:

LOCUS BI769604 846 bp mRNA linear EST 25-SEP-2001  
 DEFINITION 60305485F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5204469 5',

mRNA sequence.

ACCESSION BI769604 GI:15761182

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 846)

NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM1512 row: j column: 22

High quality sequence stop: 838.

Location/Qualifiers

1..846

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5204469"

/clone\_lib="NIH\_MGC\_122"

/lab\_host="DH10B"

/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;

Site\_1: Not; Site\_2: EcoRV (destroyed); RNA source

anonymous pool of 24 week female lung, 16 week female

spleen, and 20-22 week male spleen; Library is oligo-dT

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.4 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 026. Note:

this is a NIH\_MGC Library."

BASE COUNT 179 a 240 c 226 g 201 t

ORIGIN

alignment\_scores:

Quality: 1374.00 Length: 293

Ratio: 4.907 Gaps: 3

Percent Similarity: 95.563 Percent Identity: 95.563

alignment\_block:

US-09-357-675C-25 x BI769604 ..

Align seg 1/1 to: BI769604 from: 1 to: 846

7 LeuAlaProAspArgProAspArgThrLeuArgMetValLeuAla1 23

|||||

3 CTGGCTCAGACGCGCCCTCGGATCGGACCGCTGGATGGTTGGCTAT 52

|||||

23 eSerSerCysArgThrTySerLeuSerAlaArgArgProArgLeuGlyPhe1 40

|||||

53 ATCTTCATGCTG.....GGCTTCA 71

|||||

40 LeThrArgProPheAlaArgPheLeuSerLeuLeuCysProGlyLeuArg 56

|||||

72 TCACAGGCGCCCTTCACAGATTCTGTGCTTCGTGTCCGGACTCCG 121

|||||

57 IleProGlnLeuSerValLeuCysAlaGlnProArgProAlaGlnMetAl 73

|||||

122 ATACCTCAACTCTCAGTACTTGTGCTCAGCCAGGCCAGAGGCATGCG 171

|||||

73 aIleSerSerSerCysGlnLeuProLeuValAlaValCysGlnValT 90

|||||

172 TATCTCCCTCTCTCTGCGAAGTGCCTGTGCTGTGCTGTGCGAGTAA 221

|||||

90 hrSerThrProAspLysGlnGlnAsnPhelystnrcysAlaGlnLeuVal 106

|||||

222 CATGAGCGCCAGACAGACAGAACCTTTAAACATGTCGAGAGCTGTT 271

|||||

107 ArgGlnAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGlnAlaPh 123

|||||

272 CGAGAGCGCTGCCAGACTGGGTGCTGCTGCTGCTGCTGCTGCTGAGGCA 321

|||||

123 eAspPheIleAlaArgAspProAlaGlnThrLeuHisLeuSerGlnProL 140

|||||

322 TGACCTTCATTCACGGGACCGCTGCAGAGAGCGCTACCGTCTGTGAACAC 371

|||||

140 eugGlyLysLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 156

|||||

372 TGGGTGGGAACTTTGGAAAGATACACCCAGCTTCCAGGGAATGTGGA 421

|||||

157 LeuThrLeuSerLeuGlyGlyPheHisGlnArgGlyGlnAspTrpGlnG 173

|||||

422 CTCTGGCTGCTCTTGGGTGGTTCATGAGCGTGGCCAAAGCTGGAGCA 471

|||||

173 nThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAla 190

|||||

472 GACTCAGAAATCTACAACTGTACGCTGCTGTGAACAGCAAAAGGGGAG 521

|||||

190 aValAlaThrTyrArgLysThrHisLeuCysAspValGlnIleProGly 206

|||||

522 TAGTGGCCACTTACAGAGAACACATCTGTGATGAGATGATTCACAGG 571

|||||

207 GlnGlyProMetCysGlnSerAsnSerThrMetProGlyProSerLeuG 223

|||||

572 CAGGGCGCTATGTGGAAGCAACTCTACATGCTGGCCAGCTCTGGA 621

|||||

223 userProValSerThrProAlaGlyLysIleGlyLeuAlaValCysTyr 239

|||||

622 GTCACTGTCTCAGACACACAGGAGCAAGATTGCTTACCTGTCATGCTAT 671

|||||

240 AspMetArgPheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlnAlaG 256

|||||

672 GACATCGCGTTCCTCGAAGCTCTCTGTGCAATGTGCTCAAGCTGAGCAGA 721

|||||

256 uIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisT 273

|||||

722 GATACCTTACTATCCCTTCAGCTTTTGATTCATTCACAGGCCAGC. CACT 770

|||||

273 rpgIValLeuLeuAlaArgAlaAlaGlnIleGlnIleCysTyrValVal 289

|||||

771 GGGAGGTGTCTCTGGGGC. CGTGTATTCGAACCAAGTCTATGTAGTGT 819

|||||

290 AlaAlaAlaGlnCysGlyArgHis 297

|||||

820 GCAGCCAGCACAGTGTGAGCCGAC 844

|||||

seq\_name: gb\_est2:BG436916

seq documentation block:

LOCUS BG436916 793 bp mRNA linear EST 14-MAR-2001

DEFINITION 602488424F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4620892 5',

mRNA sequence.

ACCESSION BG436916

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 793)  
 AUTHORS NIH-MGC  
 TITLE NIH-MGC http://mgi.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: DCTP/DTP/Gazdar  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1000382 row: 0 column: 05  
 High quality sequence stop: 767.  
 Location/Qualifiers  
 1..793  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4620892"  
 /clone\_lib="NIH-MGC\_18"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pORF7; Site:1; XhoI; Site:2;  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GCCACGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH-MGC Library."  
 BASE COUNT 181 a 210 c 230 g 172 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 1298.00 Length: 267  
 Ratio: 4.992 Gaps: 5  
 Percent Similarity: 97.378 Percent Identity: 95.880  
 alignment\_block:  
 US-09-357-675c-25 x BG436916 ..  
 Align seg 1/1 to: BG436916 from: 1 to: 793  
 81 LeupProLeuValAlaValCysGlnValThrSerThrProAspLysGlnG 97  
 |||||  
 3 CTGCCCTGGTGGCTGTGCTGCAGGTAAACATCGCCGACAGCAACAA 52  
 |||||  
 97 naanPhelyThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 114  
 |||||  
 53 GAACCTTAAACATGCTGCTGAGCTGCTGAGAGCGCCAGACATGGGTG 102  
 |||||  
 114 lacysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPro 130  
 |||||  
 103 CCGGCCCTGGCTTCCTGCTGCAGGCAATTTGACTTCATTCAGCGGACCT 152  
 |||||  
 153 GCAGAGCGCTACACCTGCTGAAACCACTGGTGGGAACCTTTGGAAGA 202  
 |||||  
 147 uTyThrGlnLeuAlaArgGlyCysGlyLeuThrPleuSerLeuGlyP 164  
 |||||  
 203 ATACACCCACCTTCCAGGAATGTGACCTGCTGCTGCTGGGTGTT 252  
 |||||  
 164 hehisiuArgGlyGlnAspTrpGluGlnThrGlnLysIleTyraAspCys 180  
 |||||  
 253 TCCATATGAGCTGGCCAAAGACTGGGAGCAGACTCAGAAAATCTCAATTGT 302  
 |||||  
 181 HisValLeuLeuAsnSerLysGlyAlaValAlaAlaThrTyraArgLysTh 197

|||||  
 303 CACGTCTGCTGAACACAGCAAGGGCAGTAGTGCCACTTAACAGAGAGAC 352  
 |||||  
 197 rHisLeuCysAspAlaGluIleProGlyGlnGlyProMetCysGluSerA 214  
 |||||  
 353 ACATCTGTGACGTAGATGATTCAGGGCAGGGGCTATGTGTAAAGCA 402  
 |||||  
 214 snSerThrMetProGlyProSerLeuGluSerProValSerThrProAla 230  
 |||||  
 403 ACTTACCATGCTCGGGCCAGTCTTGAGTCACCTGTGCACACACAGCA 452  
 |||||  
 231 GlyLysIleGlyLeuAlaValCysTyraSpMetArgPheProGluLeu 247  
 |||||  
 453 GGCAGATGTGGTACGCTGCTGCTATGACATGCGGGTTCCTGAACTCTC 502  
 |||||  
 247 rLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyPro\_SerAla 263  
 |||||  
 503 TCTGGCATGTGCTCAACGTGAGCAGAGATACCTTATCTTCACAGCT 552  
 |||||  
 264 PheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAla\_A 280  
 |||||  
 553 TTTCGATCCATTTACAGGCCACAGCCACTGGAGAGTGTTCCTGGGGCCC 602  
 |||||  
 280 rGlaIleGluThrGlnCysTyraValAlaAlaAlaGlnCysGlyArg 296  
 |||||  
 603 GTGCTATGCAAAACCCAGCTGATGATGACAGCAGCAGATGGAGCAGC 652  
 |||||  
 297 HisHisLysArgAlaSerTyrglyHisSerMetValAlaAsp\_Prof 313  
 |||||  
 653 CACCATGAGAGAGACAGTATATGSCCAGACATGTGTGTAGACCCCT 702  
 |||||  
 313 rPgly\_ThrValAlaAlaArgCysSer\_GluGlyProGlyLeuCysLeuA 329  
 |||||  
 703 GGGGAACAGTGTGGGCCCGCTGCTCTTGAAGGGCCAGGCTTCGCTTC 752  
 |||||  
 329 laArgIleAspLeuAsnTyrlLeuArgGlnLeuArgArgHis 342  
 |||||  
 753 CCCGAATGACATCAAGTATCTGCGACAGTTCGCCCAAC 793  
 seq\_name: gb\_est2:BG703176  
 seq\_documentation\_block:  
 LOCUS BG703176 790 bp mRNA linear EST 07-MAY-2001  
 DEFINITION 602686082F1 NIH-MGC\_95 Homo sapiens cDNA clone IMAGE:4818405 5',  
 mRNA sequence.  
 ACCESSION BG703176  
 VERSION BG703176.1 GI:13975252  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 790)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE NIH-MGC  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
 Toshitsuki and Piero Carninci (RIKEN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1000382 row: d column: 22  
 High quality sequence stop: 766.  
 Location/Qualifiers  
 1..790  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4818405"

```

/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/Note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTT-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      167 a      228 c      206 g      189 t
ORIGIN

```

```

alignment_scores:
  Quality: 1287.00      Length: 272
  Ratio: 5.008          Gaps: 3
  Percent Similarity: 94.485      Percent Identity: 93.750

```

```

alignment_block:
US-09-357-675c-25 x BG703176 ..

```

```

Align seg 1/1 to: BG703176 from: 1 to: 790

```

```

2 LeuAlAlAlAlA***LeuAlAProAspArgProProAspArgThrLeuAr 18
|||||
8 CTCGCTGAGCGCGCTCTCGCTCCAGACCGCCCTCCGATGAGACCCGCG 57
18 gmetValleuAlAlleSerSerCysArgThrTrpSerLeuSerArgArgp 35
58 AATGGTTTGGCTAATCTTCATGCTG..... 84
35 roArgLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeu 51
|||||
85 .....GGCTTCATCCAGCGCTCTCTCAGAGATTCCTGCTCTCTG 126
52 CysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProAr 68
|||||
127 TGTCTGAGATCCGAGTACTCAACTCTCAGTACTTGTGCTCAGCCGAG 176
68 gProAlAlAlAlAlAlleSerSerSerCysGlnLeuProLeuValA 85
|||||
177 GCCCAAGGCGATGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 226
85 lAlAlCysGlnAlAlThrSerThrProAspLysGlnAlAlAsnPhelYsThr 101
|||||
227 CTGTGTGCCGATACATCGACGCCAGACAGACAGAACTTTAAACA 276
102 CysAlAGlnLeuValArgGlnAlAlAlAlArgLeuGlyAlaCysLeuAlAlph 118
|||||
277 TGTGCTGAGCTGGTGCAGAGGCTGCAGACTGGGCTGCTGCTGCTGCTT 326
118 eluProGlnAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 135
327 CCGCTGAGGCAATTCATTCATTCGACGGGACCTGCGAGAGCGCTAC 376
135 lAluSerGlnProLeuGlyGlyLysLeuLeuGlnGlnGlnGlnGlnGlnGln 151
377 ACCGGTCTGACACCTGGTGGGAACTTTGGAAAGATACACCCAGCTT 426
152 AlAlArgGlnCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGlnArgI 168
427 GCCAGGAAATGTGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 476
168 yGlnAspTrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 185
477 CCAAGACTGGAGCAGACTCAGAAATCTACAAATGTACATGCTGCTGA 526
185 snSerLysGlnAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 201
527 ACAGCAAGGGGCGAGTAGTGCCACTACAGGAAGACACATCTGTGTGAC 576

```

```

202 ValGlnIleProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 218
|||||
577 GTAGAGATTCACAGGCGAGGCGCTATGTGTAAGCAACTTACCAGTCC 626
218 oGlyProSerLeuGlnSerProValSerThrProAlGlyLysIleGlyL 235
|||||
627 TGGGCCAGTCTTACATCAGTCTGACACACACGACGAGGAGATTGTGTC 676
235 euAlAlCysTyr.AspMetArgPheProGlnLeuSerLeuAlAlAl 251
|||||
677 TAGCTGTCTGATTGACATGCGGCTCCCTGAACTCTCTGCGCATTTGCC 726
251 aglnAlAGlnAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 267
727 TCAAGCTGAGACAGATACCTTACCTTCCCTGAGCTTTGATCATTT 776
268 ThrGlyProAla 271
|||||
777 ACAGGCCAGGCC 788

```

```

seq_name: gb_est2:BG762506

```

```

seq_documentation_block:

```

```

LOCUS      BG762506                      764 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION 602733954F1 NIH_MGC_49 Homo sapiens CDNA clone IMAGE:4859292 5',

```

```

ACCESSION  BG762506

```

```

VERSION    BG762506.1 GI:14073159

```

```

KEYWORDS   EST.

```

```

SOURCE     human.

```

```

ORGANISM   Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

REFERENCE  1 (bases 1 to 764)

```

```

AUTHORS   NIH-MGC http://img.nci.nih.gov/.

```

```

TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)

```

```

JOURNAL    Unpublished (1999)

```

```

COMMENT    Contact: Robert Strausberg, Ph.D.

```

```

Email: cgaabs-r@mail.nih.gov

```

```

Tissue Procurement: ATCC/DCPD/PTP

```

```

CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

```

DNA Sequencing by: Incyte Genomics, Inc.

```

```

Clone distribution: MGC clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/LLNL at:

```

```

http://image.llnl.gov

```

```

plate: L1CM1714 row: 1 column: 13

```

```

High quality sequence stop: 749.

```

```

Location/Qualifiers

```

```

1..764

```

```

/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"

```

```

/clone="IMAGE:4859292"

```

```

/clone_lib="NIH_MGC_49"

```

```

/tissue_type="melanotic melanoma, high MDR (cell line)"

```

```

/lab_host="DH10B (phage-resistant)"

```

```

/Note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:

```

```

EcoRI; CDNA made by oligo-dT priming. Directionally cloned

```

```

into EcoRI/XhoI sites using the following 5' adaptor:

```

```

GGCAGAG(G). Size-selected >500bp for average insert size

```

```

1.8kb. Library constructed by Ling Hong in the laboratory

```

```

of Gerald M. Rubin (University of California, Berkeley)

```

```

using ZAP-CDNA synthesis kit (Stratagene) and Superscript

```

```

II RT (Life Technologies). Note: this is a NIH_MGC

```

```

library."

```

```

BASE COUNT      162 a      218 c      201 g      183 t
ORIGIN

```

```

alignment_scores:
  Quality: 1269.00      Length: 266
  Ratio: 5.076          Gaps: 3

```

Percent Similarity: 93.985 Percent Identity: 93.985

# Alignment\_block:

US-09-357-675C-25 x BG762506 ..

Align seg 1/1 to: BG762506 from: 1 to: 764

```

10 AsparGProProasparGThrLeuArgMetValLeuAlaIleSerGly 26
   |||||||
2  GACCGCCCTCCGAGCTGCGAAGTGTGGCTATATCTTAC 51
   |||||||
26 sArgThrTyrSerLeuSerArgArgProArgLeuGlyPheIleThrArg 43
   |||||||
52 CCG.....GGCTTCATCACCAGGC 70
   |||||||
43 roProHisArgPheLeuSerLeuLeuGlyProGlyLeuArgIleProGln 59
   |||||||
71 CTCCTACAGATTCTCTGCTCTGTGCTGTGACTCCGATACCTCTA 120
   |||||||
60 .LeuSerValLeuGlyAlaGlnProArgProArgAlaMetAlaIleSer 76
   |||||||
121 ACTCTAGACTTGTGTGCTACGCCAGGCCAGACCATGCTATCTCT 170
   |||||||
76 eSerSerGlyGluLeuProLeuValAlaValCysGlnValThrSerThr 92
   |||||||
171 CTTCCTCTCGCAACTGCCCTGTGTGTGTGCGAGTACATCATGACG 220
   |||||||
93 ProAspGlyGlnGlnAspPheThrCysAlaGlnLeuValArgGlu 109
   |||||||
221 CCAGACACAGACAGACACTTTAAACATGCTGCTGAGCTGTGAGAGGC 270
   |||||||
109 aAlaArgLeuGlyAlaCysLeuAlaPheLeuProGlnAlaPheAspPhe 126
   |||||||
271 TGGCAGACAGGGGTGCTGCTGCTTCCGCTGAGGACTTTGACTTCA 320
   |||||||
126 lAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGly 142
   |||||||
321 TTGCAGGGAGCCCTGCAGAGACCTACCTGCTGACACACAGGCTGGG 370
   |||||||
143 LysLeuLeuGluGluThrThrGlnLeuAlaArgGluCysGlyLeuPhe 159
   |||||||
371 AAACCTTGGGAGAAATACACCCAGCTTCCAGGAAATGTGACTGTGGCT 420
   |||||||
159 uSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluThrGln 176
   |||||||
421 GTCCTGGGGGTTCATGAGCGGTGCCAAGACTGGAGCAGACATCAGA 470
   |||||||
176 ySleIleTyrAsnCysHisValLeuLeuAsnSerIysGlyAlaValAla 192
   |||||||
471 AAATCTACAAATTGTCACTGCTGTAACAGCAAGGGGAGTAGTGGCC 520
   |||||||
193 ThrTyrArgLysThrHisLeuGlyAspValGlnIleProGlyGlnGly 209
   |||||||
521 ACTTACAGAGACACATCTGTGTGACGTAGAGATTCCAGGCGAGGGGCC 570
   |||||||
209 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 226
   |||||||
571 TATGTGTGAAGCACTTACCATGCTGGGCCAGCTTGAAGTCACTG 620
   |||||||
226 aIleSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArg 242
   |||||||
621 TCAGGACACAGCAGCAGCAAGATTGCTGCTGCTATGACATGACATG 670
   |||||||
243 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeu 259
   |||||||
671 TTCCTGGAATCTCTCTGCGATTTGGCTCAAGCTGGAGCAGAGTACTTAC 720
   |||||||
259 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrp 273
   |||||||
721 CATCTCTCAGCTTTTGGATTCATTACAGGCCAGGCCCACTG 764

```

seq\_name: gb\_est1.AL559163

## seq\_documentation\_block:

```

LOCUS       AL559163                861 bp    mRNA    linear    EST 16-FEB-2001
DEFINITION  AL559163 LTI_NFL008_TC2 Homo sapiens cDNA clone CS0DJ014YA17 5
prime, mRNA sequence.
ACCESSION  AL559163
VERSION    AL559163.1  GI:12904391
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 861)
AUTHORS   Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES   location/Qualifiers
            source
                1..861
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CS0DJ014YA17"
                /clone_11b="LTI_NFL008_TC2"
                /sex="male"
                /issue-type="T cells from T cell leukemia"
                /note="Vector: pcwSPORT 6; Site 1: NotI; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pcwSPORT 6
                vector. Library was normalized. Library was constructed by
                Life Technologies. Contact : Feng Liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : fliang@life.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT  176 a      229 c      225 g      201 t      30 others
ORIGIN

```

## alignment\_scores:

Quality: 1263.00 Length: 299  
Ratio: 4.730 Gaps: 5  
Percent Similarity: 89.298 Percent Identity: 86.288

## alignment\_block:

US-09-357-675C-25 x AL559163 ..

Align seg 1/1 to: AL559163 from: 1 to: 861

```

17 LeuArgMetValLeuAlaIleSerSerGlyArgThrTyrSerLeuSer 33
   |||||||
2  CTGCGAATGGTTTGGCTATATCTTCAATGCTG..... 33
   |||||||
33 gArgProArgLeuGlyPheIleThrArgProProHisArgPheLeuSer 50
   |||
34 .....GGCGTCAATCACAGGCGCTCTCAGATTCCTGCTGCC 70
   |||||||
50 euleuGlyProGlyLeuArgIleProGlnLeuSerValLeuGlyAlaGln 66
   |||||||
71 TTCTGTGTCCTGACACCGGATACCTCAACTCTCACTACTTGTGCTCAG 120
   |||||||
67 ProArgProArgAlaMetAlaIleSerSerSerCysGluLeuProle 83
   |||||||
121 CCCAGGCCAGACCAATGGCTATCTCTCTCTCTGGAAGTGGCCCT 170
   |||||||
83 uValAlaValCysGlnValThrSerThrProAspGlyGlnGlnAspPhe 100
   |||||||
171 GGGGGCTGTGTGCGAGTAACTGACGCCAGATTAAGCAATGAACTTTA 220
   |||||||
100 yThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeu 116
   |||||||

```

```

221 AACATGTCGTGAGCTGTTGAGAGGCTGCAGACACTGGTGCCTGCCTG 270
117 AlpheleupProgluaAlpheaspheilaAlargaspProAlaIgluTh 133
|||||
271 GCTTTCCTGCTGAGGCACTTGCATTCATGACAGGACCTGCAGAGAC 320
133 rleuHnlsleuSerGluPro-leuGlyGlylsleuLeuGluGluThrThr 149
|||||
321 GCTACATCTCTGACCTACTGGGTGGAAACTTTGGAAAGATACATC 370
150 GlhLeuAlaArgIu-CysGlyLeuTrpLeuSerLeuGlyGlyPheHisG 166
|||||
371 CAGCTTCAGAGGATATGTGAGCTGCTGCTGTTCTTGGGTGTTCCATG 420
166 LuArgGlyGlnAspTrpGluGlnThrGlnIysIleTyraAsnCysHisVal 182
|||||
421 AGCGTGCGYMWGACTGGGACAGACTYAGAAACTCAATTCACAGCTG 470
183 LeuLeuAsnSerIysGlyAlaValAlaIleThrTyraArgIysThrHisIe 199
|||||
471 CTGCTGACAGCAAAAGGAGGAGTAGTGCCACTTACAGAGAACACATCT 520
199 uCysAspValaGluIleProGlyGlnGlyPrometCysGluSerAsnSer 216
|||||
521 GTGTGAGCTAGAGATTCCAGGGCAGGGCCTATGTTGAAAGCAACTCTA 570
216 hrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLys 232
|||||
571 CCATGCGTGGGCCAGCTGTTGAGTCACTGTCAGCACACAGCAGCAGCAAG 620
233 IlleGlyLeuAlaValCysTyraSpMetArgPheProGluLeuSerLeuAl 249
|||||
621 ATTGGTCTAGCTGTCTGTATGACATGCGTTCCTGCACCTCTCTC 670
249 AleuAlaGluAlaGlyAlaGluIleLeuThrTyProSerAlaPheGlyS 266
|||||
671 ATTGGCTCAGAGCTGGCCAGAGAACTTACTTACTTACTTACTTACTT 720
266 erile-ThrGlyProAlaHisTrp-GluValLeuLeuArgAlaArgAlaI 282
|||||
721 CCATTACAGGCCAGCCAGCTGGCAGGTGTCTGCGGGCCGCTGCTC 770
282 leguThrGlnCysTyraValAlaAlaAlaGlnCysGlyArgHisHis 298
|||||
771 WCGMAAACCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 820
299 GluLysArgAlaSerTyraGlyHisSerMetValAlaAsp 311
|||||
821 GAGAGAGAGCAACTTATGGCCACAGCATGTTGTTGTTGTTGTTGTT 859

```

```

seq_name: gb_est2:BI752623
seq_documentation_block:
LOCUS      BI752623              746 bp      mRNA      linear      EST 25-SEP-2001
DEFINITION 603028471f1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5198800 5',
            mRNA sequence.
ACCESSION  BI752623
VERSION     BI752623.1 GI:15744201
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 746)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            JOURNAL
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLML1497 row: n column: 17

High quality sequence stop: 746.

Location/Qualifiers

1..746

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="NIH\_MGC\_114"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMV-Sport6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH\_MGC Library."

BASE COUNT 154 a 220 c 197 g 175 t

ORIGIN

alignment\_scores:

Quality: 1237.50 Length: 249

Ratio: 5.135 Gaps: 3

Percent Similarity: 96.787 Percent Identity: 96.787

alignment\_block:

US-09-357-675C-25 x BI752623 ..

Align seg 1/1 to: BI752623 from: 1 to: 746

```

7 LeuAlaProAspArgProProAspArgThrLeuArgMetValLeuAlaI 23
|||||
4 CTGGCTCCAGACCGCCCTCCGGATGAGACCTGGCAATGATTTGGCTAT 53
23 eSerSerCysArgThrTyraSerLeuSerArgArg.....P 35
54 ATCTTCATGTAGGAGACTACTCCCTATCCCTGCGGGCGGATATCCAC 103
35 ro.ArgLeuGlyPheIleThrArgProProHisArgPheLeuSerLeu 51
|||||
104 CTGGCGCTGGGCTTCATCACAGGCTCTCTCACAGATTCTCTCTCTCT 153
51 uCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro 68
|||||
154 GTGTCTGAGCTCGGATTAACCTCAACTCTCACTGACTTTGTGCTCAGCCA 203
68 rGProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuVal 84
|||||
204 GGGCCAGAGCCAGTGGCTATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 253
85 AlaValCysGlnValThrSerThrProAspIysGlnGlnAsnPheLysTh 101
|||||
254 GCTGTGTCAGAGGTAACTGACGAGCCAGACAAAGCAAGCAACTTTAAAC 303
101 rCysAlaGluLeuValArgGluAlaAlaArgPheLeuValaCysLeuAla 118
|||||
304 ATGTGCTAGCTGGTTCGAGAGGCTGCGCAGACTGGGTGCTGCTGCTG 353
118 heLeuProGluAlaAlpheaspheilaAlargaspProAlaIgluThre 134
|||||
354 TCCTGCTCAGAGGCAATTCATTCATTCAGAGGACCTGCAAGAGCGCTA 403
135 HisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluIuTyraThrGln 151
|||||
404 CACCGTCTGAACCACTGGGTGGAAACTTTGGAAGATACACCCACT 453
151 uAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArg 168
|||||

```

454 TGGCAGGAAATGGACTGTGGCTGTCTGGTGGTTCCTCCAGACGGTG 503  
168 lylgnasptpgluglnthrglnlysllethyrasnrcysHsvalleu 184  
|||||  
504 GCCAGAGCTGGAGACACTCAGAAATCTACATTTCTACGTCGTCTG 553  
185 AsnSerIysGlyAlaValAlaThrTyrArgLysThrHisLeuCys 201  
|||||  
554 AACAGCAAGAGGGCAGTAGTGGCCACTTACAGAGAACATCTGTGTGA 603  
201 pValGluIleProGlyGlnGlyPrometCysGluSerAsnSerThrMet 218  
|||||  
604 CGTAGAGATTCCAGGCGAGGCTATGTGTGAAGCAACTCTACATGC 653  
218 roGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGly 234  
|||||  
654 CTGGGCCCACTCTTGAGTCACTGTACAGACACAGCAGGCAAGATAGT 703  
235 LeuAlaValCysTyrAspMetArg.PheProGluLeuSerLeu 248  
|||||  
704 CTAGCTGTCTGTATGACATGCGGTTCCTGAACTCTCTCTG 746

seq\_name: gb\_est2:BI755493

seq\_documentation\_block:  
LOCUS BI755493 813 bp mRNA linear EST 25-SEP-2001  
DEFINITION 603027349p1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:5197716 5',  
mRNA sequence.  
ACCESSION BI755493  
VERSION BI755493.1 GI:15747071  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 813)  
AUTHORS NIH-MGC htp://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bsh-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLM11495 row: a column: 13  
High quality sequence stop: 745.

FEATURES  
source  
1..813  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="5197716"  
/clone\_lib="NIH\_MGC\_114"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: EcoRV (destroyed); RNA source anonymous pool of 6  
male brains, age range 23-27 yo. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.5 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 019. Note:  
this is a NIH\_MGC Library."

BASE COUNT 155 a 241 c 225 g 192 t  
ORIGIN

alignment\_scores:  
Quality: 1223.50 Length: 278  
Ratio: 4.817 Gaps: 3

Percent Similarity: 91.367 Percent Identity: 89.209

alignment\_block:  
US-09-357-675c-25 x BI755493 ..

Align seg 1/1 to: BI755493 from: 1 to: 813

1 ProLeuAlaAlaAla\*\*LeuAlaProAspArgProProAspArgThrLe 17  
|||||  
13 CCATCGCTGGGGGCTTCTGTGCTCCAGACCGCCCTCGGATGGACCT 62  
17 uArgMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 34  
|||||  
63 GCGAATGTTTGGCTATATCTTCATGCTG..... 92  
34 rgrProArgLeuGluPheIleThrArgProProHisArgPheLeuSerLeu 50  
|||||  
93 .....GGCTTCATCACCGAGCCCTCTCCACAAATTCCTGCTT 131  
51 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67  
|||||  
132 CTGTGTCTGTGACTCCGATACCTCAACTCTCACTACTTGTGCTCAGCC 181  
67 oArgProArgAlaMetAlaIleSerSerSerSerSerCysGluLeuProLeu 84  
|||||  
182 CAGGCCAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 231  
84 AlAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLys 100  
|||||  
232 TGCGTGTGTGCGAGTAACTCAAGCCAGACAGCAAGCAAGCAACTTAA 281  
101 ThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAl 117  
|||||  
282 ACATGTCTGTAGCTGTTCGAGAGCTGCAGACTGGGTGCTGCTGCTGCTG 331  
117 aPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrL 134  
|||||  
332 TTTCCTCTCCGAGGCAATTCATTCATTCATTCATTCATTCATTCATTC 381  
134 euHisLeuSerGluProLeuGlyGlyLysLeuLeuGlnGluTyrThrGln 150  
|||||  
382 TACACTGTCTGAACCACTGGGTGGGAACTTTTGGAGAAATACACCCAG 431  
151 LeuAlaArgGlyCysGlyLeuTyrPleuSerLeuGlyGlyPheHisGluAr 167  
|||||  
432 CTTCGCGAGGAATGTGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481  
167 gGlyGlnAspTpgluglnthrglnlysllethyrasnrcysHsvalleu 184  
|||||  
482 TGGCCAAAGACTGGAGCAGACTCAGAAATCTACAAATTTGCACGTGCTGC 531  
184 euAsnSerLysGlyAlaValAlaValAlaThrTyrArgLysThrHisLeuCys 200  
|||||  
532 TGAACAGACACCGGGGCGAGTAGTGCCACTTACAGAGAACATCTGTGT 581  
201 AspValGluIleProGlyGlnGlyPrometCysGluSerAsnSerThrMe 217  
|||||  
582 GAGTAGAGATTCCAGGCGAGGCTATGTGTGCAAGCAACTCTACCAT 631  
217 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 234  
|||||  
632 GCTGTGGCCCAAGTCTTGAATCACCCTGCACGCGAGCAGCGAG.ATTG 680  
234 lylgnasptpgluglnthrglnlysllethyrasnrcysHsvalleu 250  
|||||  
681 GGCTAGCTGTCTGTATGAGCTGGGTCCCTGAACTCTCTGTGGCATTTG 730  
251 Alagln..AlaGlyAlaGluIleLeuThrTyrProSerAlaPhe..... 264  
|||||  
731 GCTCAGACGCTGGAGCGGAGATACCTTACCTATCTTACGCTTGGCGTCT 780  
265 .....GlySerIlethrglyPro 270  
|||||

781 TGGGGCCCGCCACTGGAGGCGGTGCTGGGGCT 813  
seq\_name: gb\_est2:BI757823  
seq\_documentation\_block:  
LOCUS BI757823 946 bp mRNA linear EST 25-SEP-2001  
DEFINITION 603030404P1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:5200560 5',  
mRNA sequence.  
ACCESSION BI757823  
VERSION BI757823.1 GI:15749401  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 946)  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM1502 row: h column: 01  
High quality sequence stop: 708.  
Location/Qualifiers  
FEATURES  
source 1..946  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5200560"  
/clone\_lib="NIH\_MGC\_114"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: pCMV-SPORT6; Site\_1: Nct1;  
Site\_2: EcoRV (destroyed); RNA source anonymous pool of 6  
male brains, age range 23-27 yo. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.5 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 019. Note:  
this is a NIH-MGC Library."  
BASE COUNT 213 a 264 c 270 g 199 t  
ORIGIN  
alignment\_scores:  
Quality: 1219.00 Length: 324  
Ratio: 4.247 Gaps: 8  
Percent Similarity: 88.580 Percent Identity: 81.790  
alignment\_block:  
US-09-357-675c-25 x BI757823 ..  
Align seg 1/1 to: BI757823 from: 1 to: 946  
3 AAlaAlaAla\*\*leuAlaProAspArgProAspArgThrIleuAargme 19  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
2 GCTGGCGGCTTCTGCTCCAGACCGCCCTCCGATCGAGACCCGCGAAT 51  
||| tValleuAlaIleSerSerCysArgThrTySerIleuSerArgArgProA 36  
1 | ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
52 GGT.TTGGCATATCTTCATGCTG..... 74  
36 rgleuglyPheIleThrArgProProHisArgPheIleuSerIleuLeuCys 52  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
75 .....GGCTTCATCACCAGCGCTCTCCACAGATTCCTGCTCTGTGT 119  
53 ProGlyIleuArgIleProGlnIleuSerValIleuLeuAlaGlnProArgpr 69

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
120 CCTGGACTCCGGATACCTCAACTCTCACTTGTGTGCTGACGCCAGGCC 169  
69 oArgAlaMetaIaIleSerSerSerSerCysGluIleuProIleuValaIav 86  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
170 CAGAGCATGGCTATCTCTCTCTCTCTCTCTGGAACCTGGCTGGTGGGTG 219  
86 aLysGlnValThrSerThrProAspLysGlnGlnAsnPhelysThrCys 102  
220 TGTGCCAGTAACTCAACGCGCAGACAGCAAGCAAGCAACTTAAACATGT 269  
103 AlaGluIleuValArgGluAlaAlaArgIleuGlyIaCysIleuAlaPhele 119  
270 GCTAGCTGTTGAGAGGCTGCCAGACTGGGTCTGCTGCTGCTTCTCT 319  
119 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrIleuAs1 136  
320 GCCTGAGGCAATTTGATTCATTCAGCGGAGCCCTGCAGAGACGGTACACC 369  
136 euserGluProIleuGlyGlyLysIleuIleuGluIuTyThrGlnIleuA1 152  
370 TGTCTGAACCACTGGGTGGAACTTTTGGAAAGATACACCCAGCTTGGC 419  
153 ArgGluCysGlyIleuThrIleuSerIleuGlyGlyPheHisGluArgGly1 169  
420 AGGGAATGTGACTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 469  
169 nAspTrpGluGlnThrGlnIleuIleuTyAsnCysHisValIleuLeuAsn 186  
470 AGACTGGGAGAGAGACTCAGAAATCTCAAGTGCAGCTGCTGCTGAGACA 519  
186 eLysGlyAlaValValAlaThrTyArgLysThrHisIleuCysAspVal 202  
520 GCAAGGAGGCGAGTAGTGCCACTTACAGAGACACATCTGTGACACTA 569  
203 GluIleProGlyGlnGlyProMetCysGluSerIleuSerIleuMetProG1 219  
570 GAGATTCAGGAGCAG. GGGCTATGTGTGAAGCAACTTACCATGCTGG 618  
219 yProSerIleuGluSerProValSerThrProAlaGlyLysIleuGlyLeuA 236  
619 GCCAGAGCTTGAAGTACCTGTCACACACCGGAGGAGCAAGTTGTGTAG 668  
236 IavAlCysTyAspMetArgPheProGluIleuSerIleuAlaIleuAlaGln 252  
669 CTGTCTGTATGACATGCGGTTCCTGAATTCCTCGGCGAGTGCACAA 718  
253 AlaGluIleuGluIleuIleuThr.TyrProSerAlaPheGly.SerIleThr 268  
719 GCTGAGCAGAGATGATCTTACCGTATCCGTCAGCTTGGGATCAGATTACG 768  
269 GlyPro.....AlaHisTrpGluValIleuIleuArgAlaArgAla1 282  
769 GGGCCAGACACACGTGGAGAGGTGAAG.....CTGAGGGCCAGTGTGCT 812  
282 e...GluThr...GlnCysTyValValAlaAlaAlaGlnCysGlyArgH 297  
813 ATCCGAAACGCTAGAGGCTCAGATAGTGGCGCAGACACACGG..GGACCC 861  
297 is...HisGluIleuArgAlaSerTyArgIleuHisSerMetValValAsp 312  
862 ACCCTGACGAGAGAGACCCAGATTAAGGCGCACAGCATGTGTGAAAAACCC 911  
313 TrpGlyThrValValAla 318  
912 ATGGGAAACAGAGATGCC 929  
seq\_name: gb\_est1:AL520767  
seq\_documentation\_block:  
LOCUS AL520767 992 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL520767 LFI.NF1004.NBC2 Homo sapiens cDNA clone CS0DB002YD05 3  
prime, mRNA sequence.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 992)	Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.	Full-length cDNA libraries and normalization unpublished (2001)	
Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.				
FEATURES	source			
1..992 Location/Qualifiers /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CSDB002YD05" /clone_id="T1_NFL004_NBC2" /sex="male" /tissue_type="neuroblastoma cells" /lab_host="DH10B" /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng liang life technologies, a division of invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"				
BASE COUNT	229 a	267 c	268 g	222 t
ORIGIN	6 others			
alignment_scores:				
Quality:	1201.50	Length:		236
Ratio:	5.179	Gaps:		3
Percent Similarity:	98.305	Percent Identity:		96.610
alignment_block: US-09-357-675C-25 x AL520767/rev ..				
Align seg 1/1 to reverse of: AL520767 from: 1 to: 992				
129	AspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuLeu	145		
983	GACCCGCGAGACGCTACACCTGCTGGAACCACTGGGGAAC...	937		
145	uGluGlu TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu	161		
936	TTCGRGRATACACCGCTTGGCAGGGAATGTGACTGCTGCTCTTG	887		
162	GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnIleu	178		
886	GGGGGTTTCATGAGCGTGGCCAGAGCTGGAGCAGACACAGAAATCTA	837		
176	rsnCySHisValLeuLeuAsnSerLysGlyAlaValAlaIleThrTyrA	195		
836	CAATTCCTACGCTGCTGTACAGCAAGGGGCGAGTACGACCTTACA	787		
195	rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys	211		
786	GGAGAGCACATCTGTGTGACGTAGAGATTCCAGGGCGGGCCCTATGTCT	737		
212	GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSer	228		
736	GAAAGCACTTACCATGCTGGGCCCACTTTGAGTCACTGTCTGACAC	687		
228	rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG	245		

```

|||||
686 ACCAGCAGGCAGATGGCTCTAGCTCTCGTATGATGACGCGTTCCCG 637
|||||
245 luleuSerleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuThrTyro 261
|||||
636 AACTCTCTGTGGCATTTGGCTCAAGCTGAGCAGAGATCTACTATCTCT 587
|||||
262 SerAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleu 278
|||||
586 TCAGCTTTGGATCCATTCACAGGCCAGCCACCTGGAGCTGTGTGCG 537
|||||
278 gAlaATgAlaIleuThrGlnCysTyrValValAlaAlaIleuAlaIleu 294
|||||
536 GGGCCCTGCTATCGAAACCCAGTGCATATGATGCTCCAGCAGCAGTGT 487
|||||
295 GJATGTHIshISGIdulysArGAlaSerTyrgIyHISSerMetValValas 311
|||||
486 GGACGCCACCATGAGAGAGACCAAGTATGAGCAGCATGTGTGTAGA 437
|||||
311 proTrrpGlyThrValValAlaIleuAlaIleuAlaIleuAlaIleuAlaIleu 328
|||||
436 CCCCCTGGGAGACAGTGTGGCCCGCTGCTGAGGGGCGAGCCCTGTGCC 387
|||||
328 euAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleu 344
|||||
386 TTGCGCCGATAGACCTCAACTATCTGCGCAGTGGCGCCGACACTGTGCT 337
|||||
345 ValPheGlnHISArGArProAsPleuTyrgIyAsnIleuGlyHISProle 361
|||||
336 GTGTTCACACACCGCAGGCTGACCTGATGGCAATCTGTGCACCCACT 287
|||||
361 user 362
|||||
286 GTCCT 283

seq_name: gb_est1:AL522373

seq_documentation_block:
LOCUS AL522373 960 bp mRNA linear EST 13-FEB-2001
DEFINITION AL522373 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB008YM02 3
prime, mRNA sequence.
ACCESSION AL522373
VERSION AL522373.1 GI:12785866
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 960)
Li, W. B., Gruber, C., Jesssee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seque@genoscope.cns.fr, web : www.genoscope.cns.fr.
COMMENT
Location/Qualifiers
1..960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB008YM02"
/clone_id="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain. Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-01yo(GT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng liang life
Technologies, a division of Invitrogen 9800 Molecular
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610

```



8371 Email : fliang@life.uchicago.edu  
http://fulllength.invitrogen.com"  
BASE COUNT 225 a 259 c 235 g 217 t 4 others  
ORIGIN

alignment\_scores:  
Quality: 1179.50 Length: 231  
Ratio: 5.151 Gaps: 2  
Percent Similarity: 99.134 Percent Identity: 98.268

alignment\_block:  
US-09-357-675c-25 x AL522373/rev ..

Align seg 1/1 to reverse of: AL522373 from: 1 to: 960

```
133 ThrLeuHisLeuSerGluProLeuGlyLysLeuLeu.GluGluTyrT 149
|||||
959 ACGCTACACCTGCTGTGACCACTGGGTGGAAC...TTTGAGGAAATACA 913
|||||
149 hrcLleuAlaArgGluCysGlyLeuTrieuSerLeuGlyGlyPheHis 165
|||||
912 CCCAGCTTGGCCAGGGAATGTGACTGTGCTTGGGTGGTTCAT 863
|||||
166 GluArgGlyGlnAspTrrpGluInThrGlnLysIleTyrAsnCysHisVa 182
|||||
862 GACCGTGCCCAAGACTGGGAGCAGACTCAGAAATCTACAAATTGTCACT 813
|||||
182 LleuLeuAsnSerLysGlyAlaValAlaIleThrTyrArgLysThrHisL 199
|||||
812 GCTGTGGAACAGCAAGGGGCGAGTGGCCACTTACAGGAGACATC 763
|||||
199 euCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 215
|||||
762 TGTGTGACGTAGAGATTCCAGGGCGAGGCGCTATGTGTGAAGCACTCT 713
|||||
216 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLy 232
|||||
712 ACCATGCTGGGCCAGCTTGTGAGTACCTGTGACACACACGAGGCAA 663
|||||
232 sileGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuA 249
|||||
662 GATTGGTCTAGCTGTGCTGATGACATGCG.TTCCCTGAACCTCTCTGG 614
|||||
249 laLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 265
|||||
613 CATTTGGCTCAAGCTGGAGCAGAGATACTTACTTACTTCTCAGCTTTGA 564
|||||
266 SerIleThrGlyProAlaHisTrrpGluValLeuLeuArgAlaAlaI 282
|||||
563 TCCATTACAGGCCAGCCAGCTGGAGGTGTGGCGGCCCGCTGTAT 514
|||||
282 egluThrGlnCysTyrValAlaAlaAlaGlnCysGlyArgHisHisG 299
|||||
513 CGAAACCCAGTGTATGTAGTGGCAGCAGCAGTGTGGAGCCACCATG 464
|||||
299 luLysArgAlaSerTyrGlyHisSerMetValAlaAspTrrpGlyThr 315
|||||
463 AGAAGAGAGCAAGTATGGCCACAGCATGGTGTAGACCCCTGGGGAACA 414
|||||
316 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAs 332
|||||
413 GTGGTGCCCGGTGCTCTGAGGGGCCAGGCTTGGCTTGGCCGAAATAGA 364
|||||
332 PleuAsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisA 349
|||||
363 CCTCAACTATCTCGACAGAGTTGGCCGACACACTGCTGTGTCCAGCACC 314
|||||
349 rgaRrpOAspLeuTyrGlyAsnLeuGlyHisProLeuSer 362
|||||
313 GCAGGCTGACCTCTATGCAATCTGGGTGACCCACTGTCT 273
```

THIS PAGE BLANK (USPTO)

OM of: US-09-357-675C-25 to: Issued\_Patents\_NA: \* out\_format : pfs  
Date: Apr 29, 2002 10:03 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=framed.p2n.model -DEV=xlh  
-O/cgn2\_1/USPFO.spool/US09357675/rnalt\_29042002\_091757\_3310/app\_query.fasta.1.1208  
-Db-Issued\_Patents\_NA -QPM=fasta -SUFFIX=p2n.rni  
-GADP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -OGADP=4.500 -OGAPEXT=0.050 -XGADP=10.000  
-XGAPEXT=0.500 -FGADP=6.000 -FGAPEXT=7.000 -YGADP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blsnum2 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFM=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09357675 @CGN1.1.18 -NCPu=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPRX -WAIT -THREADS=1

## Search information block:

Query: US-09-357-675C-25  
Query length: 362  
Database: Issued\_Patents\_NA: \*  
Database sequences: 383533  
Search length: 122816752  
Search time (sec): 85.140000

## Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-63 +				215.50	363.13	7.1e-13 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-63 +				215.50	363.13	7.1e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-17 +				212.50	363.50	1.5e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-25 +				214.50	367.25	9.1e-13 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-25 +				213.50	365.38	1.2e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-59 +				213.50	365.38	1.2e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-59 +				213.50	365.38	1.2e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-19 +				212.50	363.50	1.5e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-61 +				212.50	363.50	1.5e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-9 +				212.50	363.50	1.5e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-17 +				212.50	363.50	1.5e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-19 +				212.50	363.50	1.5e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-61 +				212.50	363.50	1.5e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-11 +				211.50	361.63	1.9e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-21 +				211.50	361.63	1.9e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-9 +				211.50	361.63	1.9e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-11 +				211.50	361.63	1.9e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-21 +				211.50	361.63	1.9e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-07-917-111-1 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-479-638-1 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-1 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-3 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-7 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-13 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-15 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-23 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-27 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-29 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-31 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-49 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-47 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-51 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-55 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-65 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-1 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-3 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-7 +				210.50	359.75	2.4e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-13 +				210.50	359.75	2.4e-12 1785

/cgn2\_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-15 + 210.50 359.75 2.4e-12 1785  
/cgn2\_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-23 + 210.50 359.75 2.4e-12 1785  
/cgn2\_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-27 + 210.50 359.75 2.4e-12 1785  
/cgn2\_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-29 + 210.50 359.75 2.4e-12 1785  
/cgn2\_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-31 + 210.50 359.75 2.4e-12 1785  
seq\_name: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-63

## seq\_documentation\_block:

Sequence 63, Application US/08294871A  
Patent No. 5824522

## GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Manba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiratsuki, Yoshio  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
PROCESS OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US-08/294,871A  
FILING DATE: 22-AUG-1994

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993



```

APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 PAD469
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-63

alignment_scores:
Quality: 215.50 Length: 406
Ratio: 1.094 Gaps: 17
Percent Similarity: 48.522 Percent Identity: 24.384

alignment_block:
US-09-357-675C-25 x US-08-876-398A-63 ..

Align seg 1/1 to: US-08-876-398A-63 from: 1 to: 1785

2 Leua1a1a1a1a***Leua1aProAspaArgPropoAspaArgThrLeuar 18
|||||:::|||||::: |||||:::
26 TTGCAAGACGCCGAATTCGGGTCCCGTCCGTTGACAGATCAAAAAT 75
18 gmeValLeua1a1eSerSerys..... 26
::: ||| ::||| |||
76 TTTAAGCGCTGTATGCTGCATCATATAATTTCTACTTATATGAG 125

```

[illegible]





[illegible]

```

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-294-871A-5
seq_documentation_block:
  Sequence 5, Application US/08294871A
  Patent No. 5824522
  GENERAL INFORMATION:
    APPLICANT: Ikenaka, Yasuhiro
    APPLICANT: Nakan, Hirokazu
    APPLICANT: Tanaka, Masayuki
    APPLICANT: Yajima, Kazuyoshi
    APPLICANT: Yamada, Yukio
    APPLICANT: Takahashi, Satomi
    APPLICANT: Okubo, Kazuma
    APPLICANT: Yamada, Kazuhiko
    APPLICANT: Hiraiishi, Yoshiro
    TITLE OF INVENTION: Immobilized Enzyme Preparation and
    TITLE OF INVENTION: Process for Producing D-a-Amino Acid
    NUMBER OF SEQUENCES: 70
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Foley & Lardner
      STREET: 3000 K Street, N.W.
      CITY: Washington
      STATE: D.C.
      COUNTRY: U.S.
      ZIP: 20007-5109
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentln Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/294,871A
      FILING DATE: 22-AUG-1994
      CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/971,758
      FILING DATE: 12-APR-1993
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/917,111
      FILING DATE: 07-AUG-1992
    PRIOR APPLICATION DATA:

```



```
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JM109 PAD404 (FERM BP-3913)
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-5

alignment_scores:
Quality: 213.50 Length: 406
Ratio: 1.084 Gaps: 17
Percent Similarity: 48.522 Percent Identity: 24.384

alignment_block:
US-09-357-675c-25 x US-08-294-871A-5 ..
Align seg 1/1 to: US-08-294-871A-5 from: 1 to: 1785

2 LeuAlaAlaAla**LeuAlaProAspArgProProAspArgThrLeuAr 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 TTGTCAAGCGCAAAATTCGGTTCGGTCCGGTTCACAGATCAAAAT 75
18 gmetValLeuAlaIleSerSerCys..... 26
76 TTACGCGCTTATTCGTGCTGCATGTAATATTCTACTTATGTAG 125
27 .....ArgThrTyrSerLeuSerArgArgProArgLeuGlyPhe 39
126 AATTGATTCGCCCGCGAGTACAAAGCGGTTTTCGGCGATGTGTTC 175
```

```
40 IleThr.....ArgProProHisAr 46
|||
176 ACAACGTTTCCGGCGCGTGGCGGACATCACTTGAGAAAGGACGAG 225
46 gpHeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 GTTCATGACA.....CGTCAGATGATCTTG 251
63 eucysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCys 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 CAGTGGACAACAAGTCCGATCGCGCGGAGAACCGGAACGATC 301
80 GluLeuProLeuValAlaValCysGlnValThrSerThrProAspLys 96
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 GTGCTTGTCTTTCGACATC....CTGACGAACCGCGAGCCGGG 345
96 ngInAsnPhelYsThrCysAlaGluLeuValArgGluAlaAlaArgLeu 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
346 CGCGAATTCATGCTCTCCCGAATC..... 373
113 LysAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArg 129
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 ..GGCTTACGACCTTCTCCCGCGC...TGGCATTCACCGAGAGGCC 418
130 ProAlaGlnThrLeuHisLeuSerGluProLeuGlyLeuLeuGly 146
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 GAGCTCGATACCTTCTATGAGACCGAA..ATGCCGCGCGCGGTGCC 465
146 uGluTyrThrGlnLeuAlaArgGlyCysGlyLeuTrpLeuSerLeuGly 163
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466 TCCACCTCTTGAGAAAGCGCGGAATCGGATCGGCTCATCTGAGGCT 515
163 LysPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnLys 176
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
516 ACGCTGACCTGCTGTCGAAAGCGCG.....GTCAACGCT 550
177 IleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaThr 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
551 CGCTTCACACACGCTCATTTGGTGGATAGTCAGACGAAGATCGTGGCA 600
193 rTyArgLysThrHisLeuCysAspValGluIleProGlyGlnLysProm 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 GTATCGTAAGATCCATTG.....CCGGGTACAAAG.... 631
210 eTcysGluSerAsnSerThrMetProGlyProSerLeuGluSerProVal 226
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
632 ....GAGTACGAGCGCTACCGCGCTTCAGCATCTTGAAAGCGTTAT 676
227 SerThrPro.....AlaGlyLys 232
|||
677 TTCGAGCGGGGCGATCTCGGCTCCGCTCATAGCTGACGCGCGGAA 726
232 sIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeu 249
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 AATGGGATGTTCATTCGCAAGATCGCGCTGCGCTGAAACCTCGGGGG 776
249 lAlaValAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
777 TGATGGGCTTCAGGGGCGCGAGATCATCTGC.....GGC 811
266 SerIleThrGlyProAlaHisTrpGluValLeu..... 276
812 GGTACAAACACGCGACCCACAAATCCCTTGTCCCGACGACGACACT 861
277 .....LeuArgAlaArgAlaIleGluThrG 285
862 GACGTCTTCACCATCTCTATGATGACGAGCGGGGTCTATACAGAAC 911
285 lncysTyValValAlaAlaAlaGlnCysGlyArgHisGlyLysArg 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
912 GGGCTGTGTCGCGCGCGGCAAGGTGGGCATG...GAGGAGAACTGC 958
302 AlaserTyGlyHisSerMetValValAspProTrpGlyThrValAla 318
```

```

959 ATGCTGCTCGGCACCTCTGCAATCGGCCGCGGGAATCGCGC 1008
318 aagcysserglu...glyproglyleucysleualaarglleaspleua 334
1009 TCTCACTACGACCGCTGGAAGACGAGGATCACCAGCCGCGTGTCTCG 1058
334 snTyrleuAArgGlnleuAArgHisleuProvalPhe...GlnHisArg 349
1059 ATGCGTGCCTGGCAACTGCGTGAACACATCTTCAACTTCAAGCAGCATCGT 1108
350 ArgProAspleuTyrGly 355
1109 CAGCCCGACGACATATGCT 1126

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:us-08-294-871A-59

seq_documentation_block:
; Sequence 59, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraishi, Yoshiro
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/127/AOPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; STRAIN: JMI09 pad456
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 233..1144
; US-08-294-871A-59
```

```

alignment_scores:
      Quality: 213.50      Length: 406
      Ratio: 1.084      Gaps: 17
Percent Similarity: 48.522      Percent Identity: 24.384
```

```

alignment_block:
US-09-357-675c-25 x US-08-294-871A-59
Align seg 1/1 to: US-08-294-871A-59 from: 1 to: 1785
```

```

2 leualaalaala**leualaaproaspargproproasparqthrlleuar 18
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
26 TTGTCAAGCAGCGCAAAATTCGGGTTCGCGTCCGTTGACGATCAAAAT 75
18 gmetvalleualailesercys..... 26
76 TTAAAGCCGTGTATTGTGTCGTCGATGTAATATTTCGTACTTTATGTAG 125
27 .....ArgThrTyrSerleuSerArgArgProAArgleuGlyPhe 39
126 AATTTCATTGCGCGCGAGTCACAAAGCCGTTTCGGCGATGTGTTTC 175
40 IleThr.....ArgProBHisArg 46
176 ACAACGTTTTCCCGCGCGCTGGCGCGACATCACTACAGGAAGACAGAG 225
46 gphleuSerleuLeuCysProGlyleuArgIleProGlnleuSerVal 63
226 GTTCATGACA.....CGTCAAGATGATTACTTG 251
63 eucyslaGlnProArgProArgAlaMetValIleSerSerSerCys 79
252 CAOTGGACACAAGTCCGATCGCGCGGAGACACGCGAAGAGTGC 301
80 GlnleuProleuValAlaValCysGlnValThrSerThrProAspLysG 96
302 GTGCTTCGTTCTCTCGACATG.....CTGACGAAGCCCGGCGGG 345
96 nglnAsnPhelYsThrCysAlaGlnleuValArgGlnAlaAlaArgleuG 113
```

```

346 CGGATTCATGTCCTCCCGACATC..... 373
113 lYAlaCySLeuAlaPheLeuProGluAlaPheSPHeIleAlaArgASP 129
374 ..GGCTTACAGACTTCTTCCCGCGGTAT...TTCACCGAGCAGGCC 418
130 ProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuLeuGlu 146
419 GAGCTGATAGCTTCTATGAGACCGAA..ATGCCCCGGCCGGGTCCG 465
146 uGluYrThrGlnLeuAlaArgLysGlyLeuThrPLeuSerLeuGly 163
466 TCCAGCTTGTGAGAGCGCCGGAACCTCGGATCGGCTTCAATCTGGCCT 515
163 lYPHeHis.....GluArgGlyGlnAspTrpGluThrGlnLys 176
516 ACAGTGAACCTGCTGCGAAGCGGC.....GTCAACGCT 550
177 lLeTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaThr 193
551 CGCTTCAACACGTCATTTGGTGAGTAAGTCAAGCAATCGTGGCAA 600
193 rYrYrGlySThrHisLeuCySAspValGluLeuProGlyGlnGlyProm 210
601 GATTCGTAAAGATCCATTG.....CCGGTCACAAAG... 631
210 eTcYsLysSerAsnSerThrMetProGlyProSerLeuGluSerProVal 226
632 ....GAGTACGAGCGCTACCGCCGTTCCAGCATCTTGAAGACGTTAT 676
227 SerThrPro.....AlaGlyLys 232
677 TTGAGACCGGGCGATCTCGGCTTCCGGTCTATAGCTCAGCGCGGAA 726
232 sIleGlyLeuAlaValCySTyrAspMetArgPheProGluLeuSerLeuA 249
727 AATGGGAGTGTTCATCTGCACAGATCGCGCTGGCTGAAGCCTGGGGG 776
249 lAlaValAlaGlnAlaGlyAlaGluIleLeuThrYrProSerAlaPheGly 265
777 TGATGGGCGCTCAGGGCGCGAGATCATCTGC.....GGC 811
266 SerIleThrGlyProAlaHisTrpGluValLeu..... 276
812 GGCTACACAGCGCCAGCCACAAATCCCTGTGTCGCCAGCAGCACACT 861
277 .....LeuArgAlaArgAlaIleGluThrG 285
862 GAGTCCTTCCACCATCTCTATCGATGCGAGCGGGTCTTATCAGAACG 911
285 lncYsTyValValAlaAlaGlnCySgLyArgHisShSgLyLysArg 301
912 GGCCCTGTCGCCGCGCGCGAGTCAAGCAGCATG...GAGGAGAACTGC 958
302 AlAserTyGlyHisSerMetValAlaSPTrpTrpGlyThrValAla 318
959 ATCTGCTGCGCCACTCTCATGTCGCGCCAGCCGGAACATCGTCGC 1008
318 aArgCySergLys...GlyProGlyLeuCySLeuAlaArgIleAspLeuA 334
1009 TCTCAGTACAGAGCTGGAAGACAGATGATCAGCCGCGCTCATCTCG 1058
334 snTyTrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisArg 349
1059 ATGCGTCCGCGAAGTGGTGAACACATCTTCAACTTCAAGCAGCATCGT 1108
350 ArgProAspLeuTyrgly 355
1109 CAGCCCCAGCATATGCT 1126

```

seq\_name: /sgn2\_6/ptodata/2/lna/6A\_COMB.seq:us-08-876-398A-5

```

seq_documentation_block:
; Sequence 5, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Megner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 PAD404 (FERM BP-3913)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-5

```

alignment\_scores:

Quality	213.50	Length:	406
Ratio:	1.084	Gaps:	17
Percent Similarity:	48.522	Percent Identity:	24.384

alignment\_block:

US-09-357-675C-25 x US-08-876-398A-5 ..

Align seg 1/1 to: US-08-876-398A-5 from: 1 to: 1785





```

TITLE OF INVENTION:      Immobilized Enzyme Preparation and
TITLE OF INVENTION:      Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES:    70
CORRESPONDENCE ADDRESS:
ADDRESSEE:  Foley & Lardner
STREET:    3000 K Street, N.W.
CITY:      Washington
STATE:     D. C.
COUNTRY:   U.S.
ZIP:       20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE:  Floppy disk
COMPUTER:     IBM PC compatible
OPERATING SYSTEM:  PC-DOS/MS-DOS
SOFTWARE:     PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/294,871A
FILING DATE:       22-AUG-1994
CLASSIFICATION:    435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  US 07/971,758
FILING DATE:       12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  US 07/917,111
FILING DATE:       07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  US 08/211,641
FILING DATE:       11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  JP 400848/1990
FILING DATE:       07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  PCT/JP91/01696
FILING DATE:       06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  JP 407922/1990
FILING DATE:       27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  JP 078840/1991
FILING DATE:       11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  JP 140051/1991
FILING DATE:       12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  PCT/JP92/00739
FILING DATE:       10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  JP 21692/1992
FILING DATE:       10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  PCT/JP93/01101
FILING DATE:       05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  JP 340078/1992
FILING DATE:       21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME:  Wegner, Harold C.
REGISTRATION NUMBER:  25,258
REFERENCE/DOCKET NUMBER:  74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE:  (202) 672-5300
TELEFAX:   (202) 672-5399
TELEX:     904136

INFORMATION FOR SEQ. ID NO.: 17:
SEQUENCE CHARACTERISTICS:
LENGTH:  1785 base pairs
TYPE:    nucleic acid
STRANDEDNESS:  double
TOPOLOGY:  linear
ORIGINAL SOURCE:
STRAIN:  JM109 pad434
FEATURE:
NAME/KEY:  CDS

```

[illegible]

```

226 ValSerThrPro.....AlaG1 231
      |||
674 TATTTCGACCGCGGCGATCTCGGCTCCGGGTCTATGACGTGACGCCG 723
      |||
231 YLysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerI 248
      |||
724 GAAATGGGGATGTCATCTGCACGATCGCGCTGGCTGAAGCCTGGC 773
      |||
248 euAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 264
      |||
774 GGGTGTGGGCTCAAGGGCGCGAGATCATCTGC..... 808
      |||
265 GlySerIleThrGlyProAla..... 271
      |||
809 GCGGCGCTACACACGCGCGACCATCCCGCTCCCGACGACGACCA 858
      |||
272 .....HisTrpGluValLeuLeuArgAlaArgAlaIleGluT 284
      |||
859 CCTGACGTCCTTCACCATCTCTCATGATGACGCGGCTTATCAGA 908
      |||
284 hrcGlcCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLys 300
      |||
909 ACGGGCGCTGTCGCGCGCGCGGCAAGGTGGCATG...GAGGAGAAC 955
      |||
301 ArgAlaSerTyrGlyHisSerMetValAlaAspProTrpGlyThrValVa 317
      |||
956 TGCATGCTCTGCGCGCATCTCTGCATCTGCGCGCGCGGGAATCGT 1005
      |||
317 lAlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspL 333
      |||
1006 CCTCTCACTACGACGCTGGAAGACGAGTGCATCACCGCGCGCTGATC 1055
      |||
333 euAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 348
      |||
1056 TCGATCGCTGCGCGGAATCGCGGAACACATCTTCAACTTCAAGACGAT 1105
      |||
349 ArgArgProAspLeuTyrGly 355
      |||
1106 CCGTCAAGCCCGACACTATGCT 1126

```

seq\_name: /cgn2\_6/p/tdata/2/ina/5A\_COMB.seq:US-08-294-871A-19

seq\_documentation\_block:

Sequence 19, Application US/08294871A

Patent No. 5824522

GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro

APPLICANT: Nanda, Hirokazu

APPLICANT: Takano, Masayuki

APPLICANT: Yajima, Kazuyoshi

APPLICANT: Yamada, Yukio

APPLICANT: Takahashi, Satoru

APPLICANT: Okubo, Kazuma

APPLICANT: Yamada, Kazuhiko

APPLICANT: Hiraiishi, Yoshiro

TITLE OF INVENTION: Immobilized Enzyme Preparation and Process for Producing D-a-Amino Acid

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/294.871A

```

      FILING DATE: 22-AUG-1994
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/971,758
      FILING DATE: 12-APR-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/917,111
      FILING DATE: 07-AUG-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/211,641
      FILING DATE: 11-APR-1994
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 400848/1990
      FILING DATE: 07-DEC-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP91/01696
      FILING DATE: 06-DEC-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 407922/1990
      FILING DATE: 27-DEC-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 078840/1991
      FILING DATE: 11-APR-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 140051/1991
      FILING DATE: 12-JUN-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP92/00739
      FILING DATE: 10-JUN-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 212692/1992
      FILING DATE: 10-AUG-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP93/01101
      FILING DATE: 05-AUG-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 340078/1992
      FILING DATE: 21-DEC-1992
      ATTORNEY/AGENT INFORMATION:
      NAME: Wegner, Harold C.
      REGISTRATION NUMBER: 25,258
      REFERENCE/DOCKET NUMBER: 74129/127/AOPA
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202) 672-5300
      TELEFAX: (202) 672-5399
      TELEX: 904136
      INFORMATION FOR SEQ ID NO: 19:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1785 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      ORIGINAL SOURCE:
      STRAIN: JM109 PAD435
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 233..1144
      US-08-294-871A-19

```

alignment\_scores:

Quality: 212.50 Length: 407

Ratio: 1.095 Gaps: 17

Percent Similarity: 47.666 Percent Identity: 24.816

alignment\_block:

US-09-357-675C-25 x US-08-294-871A-19 ..

Align seg 1/1 to: US-08-294-871A-19 from: 1 to: 1785

```

2 LeuAlaAlaAla***LeuAlaProAspArgProProAspArgThrLeuAr 18
|||||
26 TTGCAAGCAGCGCAATTCGCGTCCGCTCGGTTGACAGATCAAAAT 75

```









```

231 ylttstlgllylealalaglyltyrtyrpsmefargpneprogluileuSertL 248
232 ylttstlgllylealalaglyltyrtyrpsmefargpneprogluileuSertL 249
724 GAAATGGGGATGTCTATCTGTGCACAGCATCGCCGCTGGCTTAAGCTGGC 773
248 euAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyProSerAlaPhe 264
249 euAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyProSerAlaPhe 265
774 GGGTATGGGCTTCAGGCGCGCCAGATCATCTGC..... 808
265 GlySerIleThrGlyProAla..... 271
809 GGGGGCTACACAGCGCGACCCACAAATCCCTGTTCSSACAGCAGACCA 858
272 .....HisTrpGluValLeuLeuAlaArgAlaAlaIleGluT 284
859 CCGAGAGCTCTTCACCATCTCTATGCATCAGAGCCGGGTCTTATCAAA 908
284 hrcInCysTrpValValAlaAlaAlaAlaGlnCysGlyArgHisIscLys 300
909 ACGGGCGCTGTGCGCGCGCGGCGCAAGTGGGCGATG..GAGAGAAAC 955
301 ArgAlaISerTyrGlnHisSerMetValAlaISpProTrrGlyThValVa 317
956 TGCATGCTGTGTCGGCCATCTCTGCATGTCGTGGCGCGAGCGGGAATGCT 1005
317 lAlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspL 333
1006 CGCTCTCACATCAGACGCTGGAAGACGAGTGTATCACCGCCGCTGCATC 1055
333 euASnTyrLeuArgGlnLeuAlaArgHisLeuProValPhe..GlnHis 348
1056 TCGATGCTGTCCGGAGCTGGGTGAACACATCTTCACTTCAAGCAGCAT 1105
349 ArgArgProAspLeuTyrGly 355
1106 CGTCAGCCCGCAGCATYATGGT 1126

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-876-398A-19
seq_documentation_block:
Sequence 19, Application US/08876398A
Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENAKA, Yasuhiro
APPLICANT: NANBA, Hirokazu
APPLICANT: TAKANO, Masayuki
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-ARR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101

```

```

1 FILING DATE: 05-AUG-1993
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: JP 340078/1992
4 FILING DATE: 21-DEC-1992
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: JP 212692/1992
7 FILING DATE: 10-AUG-1992
8 ATTORNEY/AGENT INFORMATION:
9 NAME: Wegner, Harold C.
10 REGISTRATION NUMBER: 25,258
11 REFERENCE/DOCKET NUMBER: 74129/130
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: (202) 672-5300
14 TELEFAX: (202) 672-5399
15 INFORMATION FOR SEQ ID NO: 19:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 1785 base pairs
18 TYPE: nucleic acid
19 STRANDEDNESS: double
20 TOPOLOGY: linear
21 MOLECULE TYPE: DNA (genomic)
22 ORIGINAL SOURCE:
23 ORGANISM: Escherichia coli
24 STRAIN: JM109 PAD35
25 FEATURE:
26 NAME/KEY: CDS
27 LOCATION: join(233..1141)
28 US-08-876-398A-19
29
30 alignment_scores:
31 Quality: 212.50 Length: 407
32 Ratio: 1.095 Gaps: 17
33 Percent Similarity: 47.666 Percent Identity: 24.816
34
35 alignment block:
36 US-09-357-675C-25 x US-08-876-398A-19 ..
37
38 Align seg 1/1 to: US-08-876-398A-19 from: 1 to: 1785
39
40 2 LeuAlaIaAla***LeuAlaProspaRgProProAspArgThrLeuAr 18
41 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42 26 TTGTCAAGCAGCGGAAATTCGCGTTCCGCTCCGTTGACAGATCAAAAT 75
43 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
44 18 gmetValLeuAlaIleSerSerCys..... 26
45 76 TTTAAGCGCTGTATTATGTCGTCTGATGATTAATTTCCGACTTATGTC 125
46 .....ArgThrTyrSerLeuSerArgArgProArgLeuGlyPhe 39
47 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 126 AATTTCGATTGCGCGCGAGTCACAAACCGGTTTTCGCGCATGTGTTTC 175
49 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
50 40 IleThr.....ArgProProHisAr 46
51 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52 176 ACAACGTTTCCCGCGCTGGCCGGAACATCACCCTAAGAGAGAGAG 225
53 gPheLeuSerLeuLeuCysProGlyLeuArgTleProGlnLeuSerVal 63
54 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 226 GTTCATATACA.....CGTCGATGATACTTG 251
56 eucYsaIaGlnProArgProArgAlaMetAlaIleSerSerSerCys 79
57 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
58 252 CAGTGGAGACAAGATCCGATCGCGCGGAGACACGCAACAGATC 301
59 GluLeuProLeuValAlaValCysGlnValThrSerThrProAspLysG 96
60 ::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 302 GTCGTTGCTGCTTCGCAATG....CTGACGAAGAAGCCGCGCGGG 345
62 ngInaAspPheTyrCysAlaGluLeuValArgGluAlaIleArgLeuG 113
63 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 346 CGCGAATTTACTTCTTCCCGAATC..... 373
65
66 113 lyaIacysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsp 129

```



```
18 gheValleuAlaIleSerSerCys..... 26
   :||| :||| :||| :|||
76 TTACGCCCTGTTATTTCTCGCTGCATGTAATTTTCGTAATTATGATAG 125
27 .....ArgThrTyrSerLeuSerArgArgProArgLeuGlyPhe 39
   ||| ||| ||| |||
126 AATTTTCATTGGCCGCGAGTCAACAAAGCGGTTTCGGGATGTGTTTC 175
40 IleThr.....ArgProProHisIar 46
   ||| ||| |||
176 ACAACCTTTTCCCGCGCTGGCGGACATCATTAGGAAGAGCAGAG 225
46 gPheLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 63
   ||| :||| :||| :|||
226 GTTATGACAC.....CGTCAGATGATGTTG 251
63 eucCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCys 79
   :||| :||| :||| :|||
252 CAGTGGACACAAAGTCCGATCGCGCGGAGACACCGCAACAGATC 301
80 GlnLeuProLeuValAlaValCysGlnValThrSerThrProAspGly 96
   :||| :||| :||| :|||
302 GTGCTTCGCTTCTTCGACATG.....CTGACGAAAGCCCGAGCCGGG 345
96 ngInAsnPhelyThrCysAlaGlnLeuValArgGlnAlaAlaArgLeu 113
   ||| ||| ||| |||
346 CGCGAATTTCATTGCTCTCCCGACATC..... 373
113 lYAlaCysLeuAlaPheLeuProGlnAlaPheAspPheIleAlaArgAsp 129
   ||| :||| :||| :|||
374 ..GGCTTACAGACCTTCTCCCGCG...TGCGATTTCACGACGAGGCC 418
130 ProAlaGlnThrLeuHisLeuSerGluProLeuGlyGlyLeuLeuGly 146
   :||| :||| :||| :|||
419 GAGCTGATAGCTTCTATGAGACGAA...ATCCCGCGCGGCTGCTCG 465
146 uGluTyrThrGlnLeuAlaArgGlyLeuSerLeuPheSerLeuGly 163
   ||| ||| ||| |||
466 TCACCTCTTGAAGAGCGCGGACCTCGGATCGGCTTAATCTGGGCT 515
163 lYPheHis.....GluArgGlyGlnAspTrpGlnThrGlnLys 176
   :||| :||| :||| :|||
516 ACCGTAACTCGTCGCGAAGCGGC.....GTCACGCT 550
177 lIeTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaIarh 193
   :||| :||| :||| :|||
551 CGCTTCACACGTCATTTTGGTGAATAGTCAGCAAGATCGTCGCGCA 600
193 rTyArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProm 210
   ||| ||| ||| |||
601 GATTCGTAAATCCATTG.....CCGGGTCAACAAG... 631
210 eucCysLeuSerAsnSerThrMetProGlyProSerLeuGlnSerProVal 226
   ||| :||| :||| :|||
632 ....GAGTACGAGGCTTACCGGCGCTTCACAGCATCTTGAAGCGTAT 676
227 SerThrPro.....AlaGlyLys 232
   ||| ||| ||| |||
677 TTCGACCCCGCGGATCGGCTTCCGCTGTATAGCTGACGCGCGGAA 726
232 sIleGlyLeuAlaValCysTyrAspMetArgPheProGlnLeuSerLeu 249
   :||| :||| :||| :|||
727 AATGGGATGTTCTGTCACAGATCGCGCTGAGCCGTCGAGCCGCGG 776
249 lalaValAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 265
   :||| :||| :||| :|||
777 TGAATGGCCCTCAGGGCGCGGATCATCTGC.....GCC 811
266 SerIleThrGlyProAlaHisTrpGluValLeu..... 276
   :||| :||| :||| :|||
812 GGCTACACACGCGGACCCACATCCCGCTTCCAGCAGCAGCACCT 861
```

```
277 .....LeuArgAlaArgAlaIleGluThrG 285
   :||| :||| :||| :|||
862 GAGCTCCTTCACCATCTCCATGATGACGCGCGGCTTATCAGAACG 911
285 lncTyrValValAlaAlaAlaGlnCysGlyArgHisHisGlnLysArg 301
   :||| :||| :||| :|||
912 GGGCTGTGCGCGCGCGGCAAGGTGGGATG...GAGGAGAACTGC 958
302 AlaSerTyrGlyHisSerMetValAlaAspProTrpGlyThrValAla 318
   ||| ||| ||| |||
959 ATGCTGTGCGCGCTCTGATGCTGCGCGGCAAAATCTGTCG 1008
318 aArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLeu 334
   ||| :||| :||| :|||
1009 TCTCACTACGACGCTGCAACAGAGTATCATCCGCGCTCATCTCG 1058
334 snTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisArg 349
   :||| :||| :||| :|||
1059 ATGCTGTGCGCGGAACTGGTGACACATCTTCACTTCAACAGCATCT 1108
350 ArgProAspLeuTyrGly 355
   :||| :||| :||| :|||
1109 CAGCCCGACGACTATGTT 1126

seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:us-08-294-871A-9
seq_documentation_block:
; Sequence 9, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraishi, Yoshito
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
```



```

1059 ATGCGTCGCGGGGAGACAGCGTGAACACATCTTCACTCACTTAAGCAGACATCGT 1108
      :: |||::|||::|||::|||::||| ||| ||| ||| |||
350 ArgProAspLeuTyrGly 355
      ::|||:: |||||
1109 CAGCCCGCAGCAGCACTATGCT 1126

```

**THIS PAGE BLANK (USPTO)**